

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2004, 10:34:22 ; Search time 27.5 Seconds
(without alignments)
9883.862 Million cell updates/sec

Title: US-09-787-016a-1

Perfect score: 4689

Sequence: 1 ctccgtggccgtccgcccac.....actcttaagatcatactctg 2610

Scoring table:

BLOSOM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame-n2p.model -DEV=xlp
-Q=/cgn2_1/USFTO_spool_p/US09787016/runat_28042004_094328_26905/app_query_fasta_1.2759
-DB=SwissProt_42 -QWTF=fastan -SUFFIX=rep -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN 1.1.28 @runat_28042004_094328_26905 -NCPU=6 -ICPU=3
-NO WAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBIOCK=100 -LONGLOG
-DBVTIMEOUT=120 -WAPN -TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2770	59.1	1225	1 DAT1 HUMAN	Q9btc0 homo sapien
2	2156.5	46.0	614	1 DAT1_MOUSE	Q8c959 mus musculus
3	232	4.9	2039	1 PHF2 HUMAN	Q92576 homo sapien
4	203.5	4.3	656	1 CGBP HUMAN	Q9p0u4 homo sapien
5	194.5	4.1	660	1 CGBP_MOUSE	Q9cww7 mus musculus
6	193	4.1	483	1 CYL2_BOVIN	Q28092 bos taurus
7	189.5	4.0	705	1 TRDN_RABIT	Q28920 cryptolagus
8	188.5	4.0	1010	1 HC58_MOUSE	Q8b5g0 mus musculus
9	182.5	3.9	704	1 NP14_RAT	P41777 rattus norv
10	182	3.9	699	1 NP14 HUMAN	Q14978 homo sapien
11	175.5	3.7	2004	1 MY33 HUMAN	Q92794 homo sapien
12	174.5	3.7	1012	1 HC58 HUMAN	Q9bw10 homo sapien
13	174	3.7	700	1 TRDN_CANFA	P82179 canis famil
14	173	3.7	755	1 PREE1_HUMAN	Q92766 homo sapien
15	172	3.7	5038	1 PEE1_MOUSE	Q9qx77 mus musculus
16	170	3.6	1692	1 RIM1_HUMAN	Q8aur5 homo sapien
17	169	3.6	728	1 TRDN_HUMAN	Q13061 homo sapien
18	167.5	3.6	1087	1 NPH_MOUSE	P19246 mus musculus

19	167.5	3.6	2476	1 ATRX_MOUSE	Q61687 mus musculus
20	167.5	3.6	3644	1 MINT_MOUSE	Q62504 mus musculus
21	167	3.6	1101	1 PHF2_HUMAN	Q75151 homo sapien
22	165.5	3.5	1359	1 ATRX_CAEBL	Q9u7e0 caenorhabdi
23	163.5	3.5	1394	1 CNG4_BOVIN	Q28181 bos taurus
24	163.5	3.5	1781	1 AK12_HUMAN	Q02952 homo sapien
25	163	3.5	1096	1 PHF2_MOUSE	Q9wtu0 mus musculus
26	162.5	3.5	1805	1 NEST_RAT	P21263 rattus norv
27	161	3.4	2492	1 ATRX_HUMAN	P46100 homo sapien
28	161	3.4	2492	1 ATRX_PANTR	Q7yqm4 pan troglod
29	161	3.4	2492	1 ATRX_PONPY	Q7yqm3 pongo pygma
30	159	3.4	5085	1 PCLO_RAT	Q91ks6 rattus norv
31	158.5	3.4	679	1 RRP1_DROME	P27864 drosophila
32	157	3.3	713	1 ICAL_PIG	P12675 sus scrofa
33	156.5	3.3	778	1 TAST_HUMAN	Q12815 homo sapien
34	156	3.3	881	1 PHFE_MOUSE	Q94bh9 mus musculus
35	156	3.3	1861	1 MAP2_RAT	P15146 rattus norv
36	155.5	3.3	1234	1 JMJ_MOUSE	Q62315 mus musculus
37	155	3.3	5147	1 PCLO_HUMAN	Q9y6v0 homo sapien
38	154.5	3.3	1618	1 NEST_HUMAN	P48681 homo sapien
39	154	3.3	852	1 SRCH_RABIT	P16230 cryptolagus
40	154	3.3	1026	1 NPH_HUMAN	P12036 homo sapien
41	154	3.3	1191	1 CING_MOUSE	P59242 mus musculus
42	153.5	3.3	1338	1 ACIN_MOUSE	Q9j1x8 mus musculus
43	153.5	3.3	1560	1 SMCK_HUMAN	P41229 homo sapien
44	153.5	3.3	6632	1 UN89_CAEBL	O01761 caenorhabdi
45	153	3.3	952	1 YK15_CAEBL	P46012 caenorhabdi

ALIGNMENTS

RESULT 1

DAT1 HUMAN STANDARD; PRT: 1225 AA.

AC Q9BTC0: O15043; Q9BOW0; Q9BW03; Q9H4G6; Q9NTU8; Q9NUM8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Death associated transcription factor 1.
GN DAT1 OR KIAA0333.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Placenta;
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoi T., Raku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi N., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saitsu K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.,
RT "NEBO human cDNA sequencing project."
RL Submitted (FEB-2000) to the ENBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Garder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle B., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights K., Laird G.K., Lawlor S.,
RA Lehvaeslahti M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,

RA Marsh V.L., Martin S.L., McConachie L.J., McLeay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillips M., B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Skane C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.M., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20";
RN Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
RC TISSUE=Colon, Kidney, and Muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Blakesley R.A., Grimwood J., Schmutz J., Myers R.M.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Skalska U., Smalhus D.E.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP SEQUENCE OF 235-1225 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=9734984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Kawaguchi N., Tanaka A., Kokani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro";
RL DNA Res. 4:141-150(1997).
CC -!- FUNCTION: Putative transcription factor, weakly pro-apoptotic when
overexpressed (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus
after pro-apoptotic stimuli (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=3;
Name=1;
IsoId=Q9BTC0-1; Sequences=Displayed;
Name=2;
IsoId=Q9BTC0-2; Sequence=VSP_007208, VSP_007209, VSP_007210;
Note= No experimental confirmation available;
Name=3;
IsoId=Q9BTC0-3; Sequence=VSP_007208, VSP_007211, VSP_007212;
Note= No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous
gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC
DR EMBL; AK002127; BAA92094.1; -
DR EMBL; AL035669; CAC12755.1; ALT_SEQ.
DR EMBL; AL035669; CAC12756.1; -
DR EMBL; AL035669; CAC28883.1; -
DR EMBL; AL117379; CAB90275.1; -
DR EMBL; BC000770; AAH00770.1; -
DR EMBL; BC004237; AAH04237.1; -
DR EMBL; BC014489; AAH14489.1; -
DR EMBL; AB002331; BAA20791.1; -
DR Genew; HGNC:2680; DATF1.
DR MIM; 604140; -
DR InterPro; IPR003618; TFS2_centre.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00510; TFS2M; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Apoptosis; Nuclear protein; Zinc-finger; Alternative splicing.
FT DOVAIN 165 173 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOVAIN 185 193 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT ZN_FING 268 322 PHD-TYPE.
FT VARSPPLIC 388 423 Missing (in isoform 2 and isoform 3).
FT VARSPPLIC 566 580 /FTid=VSP_007208. CSKYLKSLHPSLIA (in
isoform 2).
FT VARSPPLIC 581 1225 /FTid=VSP_007209.
FT VARSPPLIC 566 598 Missing (in isoform 2).
FT VARSPPLIC 599 1225 /FTid=VSP_007210.
FT CONFLICT 252 252 STKEDRSSEKAAATASKTAPPGSTVGKQPA -> CMYH
FT CONFLICT 580 580 LGVLLDPSRSFWIAIPWACPGGLGVAALC (in isoform
3).
FT CONFLICT 592 592 /FTid=VSP_007211.
FT CONFLICT 1225 1225 Missing (in isoform 3).
FT CONFLICT 592 592 /FTid=VSP_007212.
FT CONFLICT 592 592 G -> E (IN REF. 1).
FT CONFLICT 592 592 T -> M (IN REF. 2).
FT CONFLICT 592 592 T -> A (IN REF. 2).
SQ SEQUENCE 1225 AA; 133023 MW; 595071E0D3CC74A2 CRC64;

Alignment Scores:
Pred. No.: 6,2e-148 Length: 1225
Score: 2770.00 Matches: 529
Percent Similarity: 93.63% Conservative: 0
Best Local Similarity: 93.63% Mismatches: 0
Query Match: 59.07% Indels: 36
DB: 1 Gaps: 1

US-09-787-016A-1 (1-2610) x DAT1_HUMAN (1-1225)

QY 265 ATGACGACGAAAGCGACCGGACCAATGAGGAGCACCTAAGGCATCAACACCCACGAGC 324
Db 1 MetAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaLysProThrSer 20

QY 325 AAAGAGTTCAGGAAACATGGGTTTTCGAAGGACCACTATCGCCACGAGAGGGCGCA 384
Db 21 LysGluPheArgLysThrTrpGlyPheArgThrThrileAlaLysArgGluGlyAla 40

QY 385 GGGACCGCGAGGCTGACCCACTGAGCGCCGCCACCCACGACGAGCAGCTGGGCTGTCC 444
Db 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60

QY 445 CTGCGCGCAGTGGGAGCGACCCACGACCTGAGCGCGTGGAGCAGTTCCTGACCAATT 504
Db 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThrile 80

QY 505 GCGCGCGCGCGCGAGAGAGCATGCTGCTCTCCCTGGAGGATTCCTGGAGGCCACG 564
Db 81 AlaArgArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100

QY 565 TCCTGCGCGCGCGCACGACGCGACGACGCTCCGAGGCGCGTGGGAAAGCGCTTCTGAG 624

Db 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
Qy 625 ACCAGAGCGGCCCGCCAGTCTGCTCCACAGCTGTGAAGAGACGACGAGCTCTCTCAA 684
Db 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
Qy 685 AAGGTGAAGAGGGGATGACACACATGACACCTCCGATAGTGACAGCGATGGCTGACC 744
Db 141 LysValLysGlyGlyAspAspHisAspAspThrSerAspSerAspSerAspGlyLeuThr 160
Qy 745 TTGAAGAGCTTCAGAAATCGCTTCGACAGGAGCGGGAACAGGAGCCCACTGAGAGGCC 804
Db 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180
Qy 805 CTGAAGAGGATCCAGAGTCGCTCGGAGAGAGCGCCGGAGAGGGTCCCGCCGAGACT 864
Db 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgArgGluGluGlyProAlaGluThr 200
Qy 865 GTGGCTCGAGGCCAGTGACACTGTGAGGCGCTCTGAGGCGCTGCGGCAAGCAGCGAG 924
Db 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
Qy 925 AACGATCAGGCGGTTGTCTCCAGGCTCGGAAAGATGACAGAGAGTAAGTTGGAGGGA 984
Db 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
Qy 985 AAGCGGCTCAGGACATCAAGATCAGAGCGCTGAGACTGGGCGGCGGACCGACCTGAA 1044
Db 241 LysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeuGlyArgProLysProGlu 260
Qy 1045 TGTGAGGTTACAGCCCAAGCCCTGATTGCAATTCGCGCCAGCTCCACAAACAAGG 1104
Db 261 CysGluGlyTyArgProAsnAlaLeuTyCysIleCysArgGlnProHisAsnArg 280
Qy 1105 TTTATGATTGCTGACCGCTGTGAAGATGGTTTCATGCGGATGTGTGGCGAATTTCT 1164
Db 281 PheMetIleCysCysAspArgCysGluGluThrPheHisGlyAspCysValGlyLys 300
Qy 1165 GAGGCTCAGAGGAGCTTTTGGAGAGGAATGGGAAGACTATATCTGCCCAACTGCACC 1224
Db 301 GluAlaAspGlyArgLeuLeuGluArgAsnGlyGluAspThrIleCysProAsnCysThr 320
Qy 1225 ATTCTGCAAGTCCAGGATGAGACTCATTTCAGAAACGGCAGATCAGCAGAGCTAAATGG 1284
Db 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
Qy 1285 AGACTGAGATGCTGATGGCAGCCGATTCACAGTATAGGACATAGACAGCAAGTCT 1344
Db 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
Qy 1345 AGCGAAGACCAAGGATAAAGGTAGAATTGAGAAAGCTGCAAAATCCAAAGTGGCAAGAG 1404
Db 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
Qy 1405 AAATCAAGATCTCCAGCTT----- 1425
Db 381 LysLeuLysIlePheGlnProGlyProGlyProValProThrGlnLeuProValLeuTrp 400
Qy 1425 ----- 1425
Db 401 GlnValLeuGluIleAlaValSerArgSerIleSerAlaPheThrLeuLeuHisCysIle 420
Qy 1426 -----GTGATAGAGCGCGCTGTGCTCCAAATGATATGCGCCCGGTGCTGTCAC 1476
Db 421 SerCysLysValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHis 440
Qy 1477 GTGGCGGACCGGCTCGGTGTACTGAGTAACTGCTATCTCCTCAACACCGCCGAGCG 1536
Db 441 ValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeuLysHisAlaAlaAla 460
Qy 1537 ACAATGAAGTTCTTAAGCTCAGGTAAAGACAGAGCCCAAGCCCTAAAGAAAAGATGAAG 1596
Db 461 ThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetLys 480

Qy 1597 ATGAAGCCAGAGAGCCCACTCTTCCGAAATGCGGTCTCTCAGGAGGTATTAAATCTCT 1656
Db 481 MetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIleSer 500
Qy 1657 TCTGTGCACAGACAGACAGCTCCAGAAAAAAGAGACCACTGAGAGGCGGCTGTGG 1716
Db 501 SerValHisLysArgProAlaProGluLysGlnThrValLysLysAlaValVal 520
Qy 1717 GTCCCTCGCGGAGTGAAGCACTCGGGAAGCAAGCAGCTTGTGAGAGCAGCAGCCGTCG 1776
Db 521 ValProAlaArgSerGluAlaLeuGlyLysGluAlaCysGluSerSerThrProSer 540
Qy 1777 TGGCGAGCGATCAATTAATGAGTAAAGCCAGAAAGACTGCTGCTCCCTCCGCG 1836
Db 541 TrpAlaSerAspHisAsnTyAsnAlaValLysProGluLysThrAlaAlaProSerPro 560
Qy 1837 TCACTGTTGTATAA 1851
Db 561 SerLeuLeuTyLys 565

RESULT 2
DAT1 MOUSE
ID DAT1 MOUSE STANDARD; PRT; 614 AA.
AC O8C9B9; O8BRG2; O9WV00;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Death associated transcription factor 1 (Death inducer-obliterato-1)
DE (DIO-1).
GN DATF1 OR DIO1 OR DIO1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Pre-B cell;
RX MEDLINE=93324176; PubMed=10393935;
RA Garcia-Domingo D., Leonardo E., Grandien A., Martinez P., Albar J.P.,
RA Izpisua-Belmonte J.-C., Martinez-A C.;
RT "DIO-1 is a novel gene involved in onset of apoptosis in vitro, whose
RT misexpression disrupts limb development."
RL Proc. Natl. Acad. Sci. U.S.A. 96:7992-7997(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo, and Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldairelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pallai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zvolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,

RA Birney B., Hayashizaki Y.;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL 60,770 full-length cDNAs";
 CC Nature 420:563-573(2002).
 CC -!- FUNCTION: Putative transcription factor, weakly pro-apoptotic when
 CC overexpressed.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; translocates to the nucleus
 CC after pro-apoptotic stimuli.
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Expressed at intermediate levels.
 CC -!- INDUCTION: Upregulated during apoptosis.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ238332; CAB48401.1; --
 DR EMBL: AK042474; BAC31270.1; --
 DR EMBL: AK044919; BAC32141.1; --
 DR MGD: MGI:1344352; Dtd01.
 DR GO: GO:0005634; C:nucleus; IDA.
 DR GO: GO:0006915; P:apoptosis; IDA.
 DR InterPro: IPR001965; Znf_PHD.
 DR Pfam: PF00628; PHD; 1.
 DR SMART: SM00249; PHD; 1.
 DR PROSITE: PS01359; ZF_PHD_1; 1.
 DR PROSITE: PS00016; ZF_PHD_2; 1.
 DR Apoptosis; Nuclear protein; Zinc-finger.
 KW DOMAIN 162 170 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 182 190 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT ZN FING 265 319 PHD-TYPE.
 FT CONFLICT 45 45 V -> A (IN REF. 1).
 FT CONFLICT 331 331 D -> N (IN REF. 1).
 FT CONFLICT 353 353 I -> V (IN REF. 1).
 SQ SEQUENCE 614 AA; 67409 MW; 15ACC5A4BAEF3539 CRC64;

Alignment Scores:

Pred. No.:	1,28e-113	Length:	614
Score:	2156.50	Matches:	450
Percent Similarity:	76.80%	Conservative:	40
Best Local Similarity:	70.53%	Mismatches:	113
Query Match:	45.99%	Indels:	36
DB:	1	Gaps:	12

US-09-787-016a-1 (1-2610) x DAT1_MOUSE (1-614)

QY	265	ATGAGCAGCAAGCGCAGCCGACCAATGAGAGCGCACCTAAGCCCATCAACCCACGAGC	324
DB	1	MetAspAspGlyGlyHisLeuSerAsnGluGluAlaProLysAlaIleLysProThrSer	20
QY	325	AAAGATTTCAGAAACATGGGTTTTCGAAAGCACCACTATCCAGCGAGCGGCGCA	384
DB	21	LysGluPheArgLysThrTrpGlyPheArgThrThrIleAlaLysArgGluGlyAla	40
QY	385	GGGAGCGCGAGGCTGACCCACTGGAGCGCCACCCACGAGCAGCTGGCGCTGCC	444
DB	41	GlyAspThrGluValAspProSerGluGlnGlnPro-----GlnGlnHisAsnLeuSer	58
QY	445	CTGGCGCGAGTGGGAGCGCCAGCGCACTGAGCGCGGTGGAGCAGTTCCTGACCAT	504
DB	59	LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGluPheLeuThrThr	78
QY	505	GGCGGGCGCGCGGAGGAGCAGCTGCTCTCCCTGGAGGATTCGTGGAGCGCCAG	564
DB	79	ValArgArgArgGlyLysLysAsnValProValSerLeuGluAspSerSerGluProThr	98
QY	565	TCCTGCGCCCGCCAGACGCGGAGACGCTCCGAGCGGCGAGCGCTTCTGAG	624
DB	99	SerSerThrValThrAspValGluThrAlaSerGluGlySerValGluSerSerSerGlu	118

QY	625	ACCAGAGCGGCCCCAGCTGCTTCCACAGCTGTGAAGGAACGACCGACCTCTTCTGAA	684
DB	119	IleArgSerGlyProValSerAspSerLeuGly--LysGluHisProAlaSerSerGlu	137
QY	685	AAGGTGAAGAGGAGGATGACCAAGTACACCTCCGATAGTACAGAGGAGGAGGAGT	744
DB	138	LysAlaLysGlyGluGluGluAspThrSerAspSerAspSerAspGlyLeuThr	157
QY	745	TTGAAAGAGCTTCAGAAATCGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGG	804
DB	158	LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProValGluArgSer	177
QY	805	CTGAAGGAGTTCAGAGTCGCTCGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGT	864
DB	178	LeuArgGlySerGlnAsnArgLeuArgLysLysArgGluGluAspSerAlaGluThr	197
QY	865	GTGGCTCCGAGGCGAGTACACTGTGAGGCGCTCTCTGCC--AGTAAGCAGGAGGCC	921
DB	198	--GlySerValGlnIleGlySerAlaGlnAspArgProLeuLysGlnGluPro	216
QY	922	GAGAACGATCAGGGGGTGTCTCCAGCTCGGGAAGAGTACAGAGAGAGTAAAGTGGAG	981
DB	217	GluAlaSerGlnGlyProValSerGlnSerGlnSerGlnSerGlnSerGlnSerGln	236
QY	982	CGAAGGCGGCTCAGGACATCAAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1041
DB	237	GlyLysAlaThrGlnGlyAsnThrGluGluAsnProArgGluAlaGlyLysProLysPro	256
QY	1042	GAATGTGAGGGTTACGACCCCAACGCTGTATTGCAATTCGCGCCAGCTCACAAAC	1101
DB	257	GluCysGluValTyrAspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsn	276
QY	1102	AGGTTTATGATTCGTGAGCGCTGCAAGAGTGTTCATGCGGATTCGTGGGCAAT	1161
DB	277	ArgPheMetIleCysCysAspArgCysGluGlnThrPheHisGlyAspCysValGly	296
QY	1162	TCTGAGGCTCGAGGAGGCTTTTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1221
DB	297	SerGluAlaArgGlyArgLeuGluGluArgAsnGlyGluAspTyrIleCysProAsn	316
QY	1222	ACCATTCTGAGTCAGGATGAGATCTATTCAGAACCGGACGATCAGCAGAGCACTAA	1281
DB	317	ThrIleLeuGlnValGlnAspGluThrAsnGlySerAlaThrAspGluGlnAspSer	336
QY	1282	TGAGACCTGGAGATGCTGAGCGACCGATGTCAAGTATAGGAACAATAGAGCAGAG	1341
DB	337	CysArgSerValGlyAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGln	356
QY	1342	TCTAGCGAAGACCAAGGATAAAGGTTAGAAAGCTGCAATTCAGTGGCAG	1401
DB	357	SerGlyGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGly	376
QY	1402	TAGAACTCAAGATCTCCAGCTGTGATAGAGGCGCTGGTGCTCAAAATGTTATG	1461
DB	377	LysLysLeuLysIlePheGlnProValGluAlaProGlyAlaProLysCysIleGly	396
QY	1462	CCCGGGTGTCTACGTGGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1521
DB	397	ProGlyCysSerSerValAlaGlnProAspSerValTyrCysSerAsnAspCysIle	416
QY	1522	AAACGCGCGCAGCAGCATGAGTTTCTAGCTCAGTAAAGAACAGAGCAAGCCT	1581
DB	417	LysHisAlaAlaAlaThrMetArgPheLeuSerGlyLysGluGlnLysThrLysPro	436
QY	1582	AAAGAAAGATGAGATGAAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1641
DB	437	LysGluLysValLysThrLysProGluLysPheSerLeuProLysCysSerValGln	456
QY	1642	GATTTAAATCTCTCTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1701
DB	457	GlyIleLysIleSerSerValHisLysArgLeuAlaSerGluLysArgGluAsnPro	476

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF091622; R4P21292.1; -.
 CC EMBL; D87685; BAAL3438.2; ALT_INIT.
 CC EMBL; AL050329; CAB99358.1; -.
 CC Genbank; HGNC:8921; PHF3.
 CC MIM; 607789; -.
 CC InterPro; IPR003618; TFS2 centre.
 CC InterPro; IPR001965; Znf_PHD.
 CC Pfam; PF06628; PHD; 1.
 CC SMART; SM00249; PHD; 1.
 CC SMART; SM00510; TFS2M; 1.
 CC PROSITE; PS01359; ZF_PHD_1; 1.
 CC PROSITE; PS00016; ZF_PHD_2; 1.
 CC Zinc-finger. 717 772 PHD-TYPE.
 FT ZN FING 717 772
 FT DOMAIN 929 1029 TFS2M.
 FT DOMAIN 1797 1865 PRO-RICH.
 SQ SEQUENCE 2039 AA; 229480 MW; FDAFF00576005E9B CRC64;

Alignment Scores:
 Pred. No.: 9, 15e-06 Length: 2039
 Score: 232.00 Matches: 94
 Percent Similarity: 36.43% Conservative: 63
 Best Local Similarity: 21.81% Mismatches: 154
 Query Match: 4.95% Indels: 120
 DB: 1 Gaps: 17

US-09-787-016a-1 (1-2610) x PHF3_HUMAN (1-2039)

QY 538 TCCCTGGAGGAT---TCTGGTGGAGCCACCTCTGCGCCGACACAGCCGAGAGAGCC 594
 Db 595 SerLeuSerAspLysSerHisAlaHisProGlyCysLeuLysGluProHisHisProAla 614
 QY 595 TCCGAGGCGAGCGTGGAAAGCGCTTCTGAGACACAGA---AGCGGCCCCAGCTCTGCTCC 651
 Db 615 GlnThrGlyHisValSerHisSerSerGlnCysHisLysProGlnGlnAla 634
 QY 652 ACAGCTGTGAAGAGACGACCCCTCTTCTGAAAGGTGAA---GGAGGGGATGACAC 708
 Db 635 ProAlaMetLysThrAsnSerHisValLysGluGluLeuHisProGlyValGluHis 654
 QY 709 GATGACACCTCGATAGTGACAGCGATGCGCTTGAAGAG---CTT 756
 Db 655 -----PhelLysGluGluAspLysLeuLysLysProGluLysAsnLeu 670
 QY 757 CAGATTCGCTTCGAGGAGGCGGAGGAGCCCTGAGAGCCCTGAGAGCCCTGAAAGGATC 816
 Db 671 GlnProArgGlnArgSerSerLysSerPheSerLeuAspLysProLeuPheLeu 690
 QY 817 CAGACTCGCTCGGAGAGAGCGCGGAGGAGGCTCCCGGAGACTGTGGGCTCCGAG 876
 Db 691 ProAspAsnIleAlaThrIleArgArgGlu-----GlySer--- 702
 QY 877 GCCAGTGCAGTGTGGAGGCGCTCTGCCAGTAAGACAGAGCCGAGAGACGATCAGGG 936
 Db 702 ----- 702
 QY 937 GTTGTGTCCAGGCTGGGAGAGATGACAGAGAGTAAAGTTCGAGGAGAGCGGCTCAG 996
 Db 703 -----AspHisSerSerPheGluSerLysTyr----- 712
 QY 997 GACATCAAGATGAGAGCGCTGGAGACTTGGGCGCCAGGAGCCCTGAAATGTGAGGTTAC 1056
 Db 713 -----MetTrpThrProSerLysGlnCys---GlyPhe 722
 QY 1057 GACCCCAAGCGCTGTATGATTTGCGCGAGCGCTCACAACAGAGGTTATGATTC 1116
 Db 723 -----CysLysLysProHisGlyAsnArgPheMetValGly 734

QY 1702 AAGAAGCGAGTGTGCTCCCTGCGCGAGTGAAGCACTCGGAGGAGAGAGAGCTTGTGAG 1761
 Db 477 LysLys---ValMetLeuAlaSerArgSerGluThrSerGlyLysGluAlaCysGlu 495
 QY 1762 AGCAGACCGCGCTGCGGAGCGATCAATTAATGCTGCTGAGGAGGAGGAGGAGGAGGAG 1821
 Db 496 SerSerThrProSerTrpAlaSerAspHisAsnSerAspAlaValLysProGluLysPro 515
 QY 1822 GCTGCTCCCTCC-----TCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1872
 Db 516 GluLysProThrAlaLeuSerProThrLeuLeuSerLysCysThrThrHisProLysAla 535
 QY 1873 GCGCT-CTTGAGCCCTCCGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1919
 Db 536 GlyPheProGlyPro-----SerHisLysLeuGlyGlyCysLeuGly 549
 QY 1920 CTGTCTCAGAGCT-----GGAGTGTGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1970
 Db 550 LeuSerArgThrArgValLeuValLeuValLeuValLeuValLeuValLeuValLeuVal 569
 QY 1971 GGTGACCATCAGCGGAG 2030
 Db 570 Ala-----ArgSerArgTyrGlnAspAlaSerGlyProGlnValPhe 583
 QY 2031 ATCCCTAGTCTGTCAGGACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2090
 Db 584 LeuProSerLeuTrpSerLeuSerGlyTrpPheLeuLysSerCys----- 598
 QY 2091 AGCTTACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2144
 Db 599 ---ValGlyLeuMetLeuGluAla---IleSerTyrPheSerPheArgProTrp 614

RESULT 3
 PHF3_HUMAN
 ID PHF3_HUMAN STANDARD; PRT; 2039 AA.
 AC Q92576; Q9N016; Q9U145;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE PHD finger protein 3.
 GN PHF3 OR KIAA0244.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
 RX MEDLINE=21845875; PubMed=11856869;
 RA Fischer U., Struss A.-K., Hemmer D., Michel A., Henn W.,
 RA Steudel W.-I., Meese E.;
 RT "PHF3 expression is frequently reduced in glioma."
 RL Cytogenet. Cell Genet. 94:131-136(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=97191544; PubMed=9039502;
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
 RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. VI.
 RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
 RT analysis of cDNA clones from cell line KG-1 and brain."
 RL DNA Res. 3:321-329(1996).
 RN [3]
 RP SEQUENCE OF 83-2039 FROM N.A.
 RA Matthews L.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Expression is significantly
 CC reduced or lost in glioblastomas, glioblastoma cell lines,
 CC anaplastic astrocytomas, and astrocytomas.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

GO: GO:001563; F:transcriptional activator activity; IDA.
DR GO: GO:0045322; F:unmethylated CpG binding; IDA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR InterPro: IPR002857; Znf_CXXC.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF06328; PHD_1.
DR Pfam: PF02008; zf-CXXC; 1.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS01359; ZF_PHD_1; 1.
DR PROSITE: PS0016; ZF_PHD_2; 1.
KW Transcription regulation; Activator; DNA-binding; Zinc-finger; Zinc;
KW Metal-binding; Coiled coil; Nuclear protein.
FT ZN_FING 28 76 PHD-TYPE
FT ZN_FING 161 208 CXXC-TYPE
FT DOMAIN 256 317 ASP/GLU-RICH (ACIDIC)
FT DOMAIN 321 359 ARG/LYS-RICH (BASIC)
FT DOMAIN 422 474 COILED COIL (POTENTIAL)
FT DOMAIN 156 161 POLY-GLN
FT DNA_BIND 162 221 D -> N (IN REF. 1).
FT CONFLICT 117 117 H -> N (IN REF. 1).
FT CONFLICT 302 302 K -> KMER (IN REF. 4; AAH15733).
FT CONFLICT 340 340 K -> KMER (IN REF. 4; AAH15733).
SQ SEQUENCE 656 AA; 75711 MW; 6D2376B49905A18 CRC64;

Alignment Scores:
Pred. No.: 0.000308 Length: 656
Score: 203.50 Matches: 82
Percent Similarity: 36.55% Conservative: 43
Best Local Similarity: 23.98% Mismatches: 98
Query Match: 4.34% Indels: 119
DB: 1 Gaps: 20

US-09-787-016A-1 (1-2610) x CGBP_HUMAN (1-656)
QY 976 TTGAGCGGAAAGCGGCTCAGACATCAAGATGAGGAGCTGGAGACTTGGCGGACCG 1035
DB 1 MetGluGly-----AspGlySerAspProGluProAspAlaGlyLeuAsp 16
QY 1036 AAGCCTGAATGTGAGGTTTACAGCCCAACCGCCTGATTGATTCGCGCCAGCCTCAC 1095
DB 17 Ser---LysSerGluAsnGlyGluAsnAlaProIleTyrCysIleCysArgLysProAsp 35
QY 1096 ACAACAGGTTTATGATTGCTGTGACCGCTGTGAGAGATGTTTCATGCGCATTTGTG 1155
DB 36 IleAsnCysPheMetIleGlyCysAspAsnGluTrpPheHisGlyAspCysIle 55
QY 1156 GGCATTCTGAGCTCGAGGGAGGCTTTTGGAAAGGAATGGGAGACATATCTGCCCA 1215
DB 56 ArgIleThrGluLysMetAlaLysAlaIle-----ArgGluTrpTyrCysArg 71
QY 1216 AACTGC-----ACCATTTCTGCAAGTGCAG----- 1239
DB 72 GluCysArgGluLysAspProLysLeuGluIleArgTyrArgHisLysLysSerArgGlu 91
QY 1240 -----GATGAGACTATTGAGAAACCGCAGATCAGCAGGAAGCTAAATGGAGA 1287
DB 92 ArgAspGlyAsnGluArgAspSerSerGluProArgAspGluGlyGlyArgLysArg 111
QY 1288 CTT---GGAGATGCTGATGCGCCCGATTTGATGAGTATAGGAACAATAGCAGAGAAGTCT 1344
DB 112 ProValProAspProAsp-----LeuGlnArgArgAla 122
QY 1345 ACGAAGACCAAGGATTAAGGGTAGAANTTGAAA---GCTGCAATCAAGTGCAG 1401
DB 123 GlySerGlyThrGlyValGlyAlaMetLeuAlaArgGlySerAlaSerProHisLysSer 142
QY 1402 AAGAACTCAAGATCTTCAGGCTGTGATAGAGGCGCCT----- 1440
DB 143 SerPro-----GlnProLeuValAlaThrProSerGlnHisGlnGln 158
QY 1441 -----GGTGCCTCAAAATATTGCGCCCGGCTGTGTCAGTGGCG 1482
DB 159 GlnGlnGlnIleLysArgSerAlaArgMetCys-----GlyGluCysGluAlaCys 175

QY 1483 CAGCCGACTCGGTGTACTGCGAGTAATGACTGTATCTCAACACGCGGACGACAATG 1542
DB 176 ArgArg-----ThrGluAspCys-----GlyHisCysAspPheCysArg 188
QY 1543 AAGTTTCTAAGCTCAGTAAAGAACAGAACCAAGCCCTAAAGAAAGATGAAG----- 1596
DB 189 AspMetLysLysPheGlyGlyProAsnLysIleArgGlnLysCysArgLeuArgGlnCys 208
QY 1597 -----ATGAGCCAGAGAAGCC 1614
DB 209 GlnLeuArgAlaArgGluSerTyrLysTyrPheProSerSerLeuSerProValThrPro 228
QY 1615 -----AGTCTTCGAAATGCGGTGCTCAGCAGGATTAAATCTCTCTGTGACACAAG 1668
DB 229 SerGluSerLeuProArg-----ProArg 236
QY 1669 AGACCACTCCAGAAAGAAAGAGACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1728
DB 237 ArgProLeuProThrGlnGlnGln-----ProGlnPro 247
QY 1729 AGTGAAGCACTCGGGAAG-----CAGCAGCTGTGTGAGCAGCAGCAGCAGCAGC 1776
DB 248 SerGlnLysLeuGlyArgIleArgGluAspGluGlyAlaValAlaSerSerThr----- 265
QY 1777 TGGCGAGCGATCAATCAATGCACTAAAG-----CCAGAAAGAGACTGCTGCTCCC 1830
DB 266 -----ValLysGluProProGluAlaThrAlaThrPro 276
QY 1831 TCGCGG 1836
DB 277 GluPro 278

RESULT 5
CGBP_MOUSE STANDARD; PRT; 660 AA.
ID CGBP_MOUSE
AC QSCW7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE CpG binding protein (Protein containing PHD finger and CXXC domain 1).
GN CGBP OR PCX1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=Embryonic stem cells;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuura Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga M., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guernicchi S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang X.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:695-690(2001)
CC -!- FUNCTION: Transcriptional activator that exhibits a unique DNA
CC binding specificity for [AC]CpG[AC] unmethylated CpG motifs (By
CC similarity).

CC -|- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -|- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -|- SIMILARITY: Contains 1 CXC-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

-----CC
EMBL; AK010337; BAB26862.1; -.
DR
DR MGD; MGI:1921572; Cgdp.
DR
DR GO; GO:0016363; C:nuclear matrix; IDA.
DR InterPro; IPR002857; Znf_CXXC.
DR
DR UniProt; P000195; Znf-⁷⁰⁵UniProt

DR tucelfr10; tfr001903; zml_fhd.
DR pfam; PF00628; PHD; 1.
DR pfam; PF02008; zf-CXXC; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.

KW	Transcription regulation; Activator; DNA-binding; zinc-finger; Zinc;
KM	Metal-binding; Coiled coil; Nuclear protein.
FT	2N_RING 28 76 PHD-TYPE
FT	2N_FING 165 212 CXXC-TYPE
FT	DOMAIN 260 321 ASP/GLU-RICH (ACIDIC).

FT	DOMAIN	325	363	ARG/LYS-RICH (BASIC).
FT	DOMAIN	426	479	COILED COIL (POTENTIAL).
FT	DOMAIN	156	161	POLY-GLN.
FT	DNA BIND	166	225	BY SIMILARITY.
SEQ	SEQUENCE	660 AA;	76166 MW;	9D92CB6DA82069EA CRC64;

Alignment Scores:	
Pred. No.:	0.000894
Score:	194.50
Length:	660
Matches:	78
Percent Similarity:	36.36%
Conservative:	46
Best Local Similarity:	22.87%
Mismatches:	104

Query Match:	4.15%	Indels:	113
DB:	1	Gaps:	18

	976	TTGACGAAAAGCGCTTACGACATCAAAAGATGAGAGCCCTGGAGACTCTGGCCGACCG	1036	AAAGCCGTAATGTGAGGTTTACGAGCCCAACGGCCCTGTATTGCACTTTGGCCGACGCTCAC
	QY	:::		
		1 MetGluGly-----AspGlySerAspLeuGluProAspAlaGly----Asp 15		
	Db			
	QY			

16	AppSerLysSerGluAsnGlyGluAsnAlaProIleTyrCysIleCysArgLysProAsp	35
Db		
1096	AAACAACAGTTTATGATTTCTGTGACCGCTGTGAGAGATGGTTTATGCGATTGTG	1155
QY		
36	IleAsnCysPheMetIleGlyCysAspAsnCysAsnGluTrpPheIleGlyAspCysIle	55
Db		

OY 1156 GGCAATTCGAGCTCGAGGAGCGTTTGGAAAGGAATGGGAGAAGACTATATCTGCCCA 1215
|||||
Db 56 ArgIleThrGlulysMetAlaLysAlaIle-----ArgGluTrpTyrCysArg 71
|||||

1216	AACTTC	-----ACCATCTTCCAAAGTCAG	-----	1237
72	GluCysArgGlnIysAspProLysLeuGluIleArgTyrzArgHisLysLysCysArgGlu	91		
Db				
1240	-----CATGAGACTCATTCAGAAAACGGCCAGATCAGCAGGAAGCTAAATGGAG	1287		
QY				

Db		92 ArgAspGlySerGlnArgAlaGlySerGluProArgAspGluGlyGlyArgLysArg	111
QY	1288	CCTGGAGTGTATGCCCGCATTTACAACTATTAGGAACAATATAGCAGCACTTAGC	1347
Db	112 Pro-----	-----AlaSerAspProGluLeuGlnArgArgAlaGly	123

QY 1348 GAAGACCAAGGATAAAGGCTAGATTGAGAA--GCTGCAATCCAAGTGCAGAAG 1404

124	SerGlyThrGlyValGlyAlaMetLeuAlaArgGlySerAlaSerProHisLysSer	143
1405	AAACTCAAGATCTTCACGCTGTGATAGAGCGCT	1440
144	Pro-----GlnProLeuValAlaThrProSerGlnHisHisGlnGlnGln	159
1441	-----CGTCCTCAAAATGATTATGGCCCGGCTGCTGT	1473
160	GlnGlnGlnGlnGlnGlnLeuLysArgSerAlaArgMetCys-----GlyGluCys	176
1474	CACGTGGCGACGCCGACTCGGTGTACTGCTAGTAATGACTGTATCTCAACACGCCGCA	1533
177	GluAlaCysArgArg-----ThrGluAspCys-----GlyHisCysAsp	189
1534	CGCAACATGAAGTTTCTAAGCTCAGGTAAAGAACAGAACCCCTAAAGAAAAAGATG	1593
190	PheCysArgAspMetLysLysPheGlyGlyProAsnLysIleArgGlnLysCysArgLeu	209
1594	AAG-----ATGAAAGCCA	1605
210	ArgGlnCysGlnLeuArgAlaArgGluSerTyrLysTyrPheProSerSerLeuSerPro	229
1606	GAGAAGCCAGT-----CTTCCGAATCGGTGCTCAGGCAGGTATTAAATCTCTTCT	1659
230	ValThrProSerGluAlaLeuProArg-----	238
1660	GTGCACAGACACCGCTCCAGAAAAAAGACACACAGTGAAGAAGCAGCTGGTGGTC	1719
239	---ProArgArgProProThrGlnGln-----	248
1720	CGTGGCGGAGTGAACACTCGGAAGGAGCAGCTGTGTGACGACGACGCGCGTGTGCG	1779
249	ProGlnGlnSerGlnLysLeuGlyArgGluAspGluGlyThrValLeuSer---	267
1780	GGGAGCGATCAATTAACATGCAGTAAG-----CCAGAAAAACAGCTCTCTCTCCCTCG	1833
268	-----SerValValLysGluProGlnGluAlaThrAlaThrProGlu	281

QY	1834	CGG	1836
Db	282	Pro	282

RESULT 6		
CYL2_BOVIN	STANDARD;	PRT; 488 AA.
ID CYL2_BOVIN		
IC 028052.		

DE Cyclicin II (Multiple-band polypeptide II).

GN
CY1C2 OR CYL2.
OS
Bos taurus (Bovine).
OS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoides;
OC
Bovidae; Bovinae; Bos

OC	Derivative, Derivative, Derivative
OX	NCBI_TaxID=9913;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Testis;

RX MEDLINE=95255491; PubMed=7737358;
 RA Hess H., Heid H., Zimbelmann R., Franke W.W.;
 RT "The protein complexity of the cytoskeleton of bovine and human sperm
 RT heads: the identification and characterization of cylicin II.";
 RT Exp Cell Res 218:174-182(1995)

RA Hess H., Heid H., Franke W.W.;
RX MEDLINE-93359502; PubMed-8354692;
RP PARTIAL SEQUENCE.
RN [2]
RT Engr. Cell Res. 210:174-202(1999).

RT "Molecular characterization of mammalian cyclin, a basic protein of
RT the sperm head cytoskeleton."; RT
RT J. Cell Biol. 123:1043-1052(1993).
CC - FUNCTION: Possible architectural role during spermatogenesis. May
CC be involved in chromatin differentiation

DE TUITIVIT IN PARVOIRIT IN SPERMATION DIFFERENTIATION.

CC -!- SUBCELLULAR LOCATION: Calyx; sperm head cytoskeletal structure.
 CC -!- TISSUE SPECIFICITY: Testis.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: Z66789; CAA86753.1; -
 CC PIR: I46014; I46014.
 DR Cytoskeleton; Structural protein; Repeat; Sperm; Spermatogenesis.
 KW DOMAIN 25 487 49 X 3 AA REPEATS OF K-K-X.
 FT DOMAIN 189 407 8 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 189 212 1.
 FT REPEAT 213 239 2.
 FT REPEAT 240 267 3.
 FT REPEAT 268 295 4.
 FT REPEAT 296 323 5.
 FT REPEAT 324 351 6.
 FT REPEAT 352 379 7.
 FT REPEAT 380 407 8.
 SQ SEQUENCE 488 AA; 53562 MW; 5P3107CE2D019C5A CRC64;
 Alignment Scores:
 Pred. No.: 0.00114 Length: 488
 Score: 193.00 Matches: 117
 Percent Similarity: 32.99% Conservative: 73
 Best Local Similarity: 20.31% Mismatches: 204
 Query Match: 4.12% Indels: 182
 DB: 1 Gaps: 19
 US-09-787-016a-1 (1-2610) x CIVL2_BOVIN (1-488)
 QY 316 CCCACGACGAGGTTCTAGGAAACATGGGGTTTTCGAGGACCACTATCCGCAAGCGA 375
 DB 19 ProValSerGluLeuSerLysLysSerTrpAsn----- 29
 QY 376 GAGGGCGCAGGGACCGGAGGCTGACCCACTGGAGCCGCCACCCACGACGAGCTG 435
 DB 30 -----GlnGlnHisphe 33
 QY 436 GGCCTCTCCCTGGCGCAGCTGGAGGAGCCCAAGCGCACTGAGCGCTGGAGCAGTTTC 495
 DB 34 AlaLeuValPheProLysProProArgProGlyLysArg----- 46
 QY 496 CTGACCATTTGGCGGCGCGCGCAGGAGAGCATGCTCTCTCCCTGGAGATTCTGTGT 555
 DB 47 -----ArgArgSerLysProSerLeuLeuGlnGlnAsn----- 57
 QY 556 GAGCCCACTCTGCGCCCGCCGACAGACGCGAG----- 588
 DB 58 -----ThrSerProLysTyrrAspAlaGluLysLeuArgGlyAspArgLysGlnPro 74
 QY 589 -----ACAGCTCCGAGGCGAGCGCTGGAAGCGCTTCT 621
 DB 75 LeuTrpMetHisArgSerLeuMetArgLleSerGluArgProSerValTyrrLeuAlaAla 94
 QY 622 GAGACCAGA-----ACGCGCCCGCAGTCTGCTTCACAGCTCTGTGAGGACGA 669
 DB 95 ArgSerArgHisProGlnLysGluThrProProSerGlnGluAspAlaLysGlnAlaAla 114
 QY 670 CCAGCCTCTTCTGAAAGGTGAAGGAGGGGATGACCCAGCATGACACCTCCGATAGTGAC 729
 DB 115 LysProSerSerProLysValLysLysSerLysGluAspLysAspLysSerAspSerGlu 134
 QY 730 ACGGATGGCTGACCTTGAAGAGCTTCAGATCCGCTTCGAGGAGCGGGAACAGGAG 789
 DB 135 AlaGluSerIleValSerLysGluLysProArgLysLeuSerAlaLysGluLys 154
 QY 790 CCCACTGAGAGG-----CCCTGAAAGGGATCCAGACT 822

DB 155 ProAspGluLysLysAspLeuLysLysGluArgLysAspSerLysLysGlyLysGluSer 174
 QY 823 -----CGCTG 828
 DB 175 AlaThrGluSerGluAspGluLysAlaGluLysGlyAlaLysLysAspArgLys 194
 QY 829 CGGAAGAAGCCCGGAGGAGGTCCGCCAGACTGTGGCTCCGAGCCAGTACACT 888
 DB 195 GlySerLysLysGlyLysGluThrProSerAspSer---GlySerGluLysGlyAspAla 213
 QY 889 GTGGAGGCGCTCTGCCCAAGTAAGCAGGAGCCCGAG-----AACGAT 930
 DB 214 LysLysAspSerLysLysSerLysLysAspSerLysGlyLysGluSerAlaThrGluSer 233
 QY 931 CAGGGGTGTGTCCAGCTGGGAAGATCACAGAGAGATTAAG----- 975
 DB 234 GluGlyGluLysGlyAspAlaLysLysAspLysLysGlyLysGlySerLysLys 253
 QY 976 -----TTGGAGGGAAGGCGCTCAGGACATCAAGATCAGGAG 1014
 DB 254 GlyLysGluSerAlaThrGluSerGluGlyLysGlyAspAlaLysLysAspLys 273
 QY 1015 CCTGAGACTTGGGCGGACCGAAG-----CCTGAATGTGAGGTTTACGAC 1059
 DB 274 LysGlyLysLysGlySerLysLysGlyLysGluSerAlaThrGluSerGluGlyLys 293
 QY 1060 CCCAAGCCCTGTATTGCTATTGCGCCGAGCTCACACACAGGTTTATGATTTGCTGT 1119
 DB 294 GlyAspAlaLysLys----- 298
 QY 1120 GACCGCTGTGAAGAATGTTTCATGGCGATGTGTGGGCATTTCTGAGGCTCGAGGAG 1179
 DB 299 -----AspAspLysLysGlyLys 304
 QY 1180 CTTTGGAAAGAAATGGGAAGACTATATCTGCCCCAACTGCCCACTTCTGCAAGTGAG 1239
 DB 305 LysGlySerLysLysGlyLysGlu-----SerAlaThrGlu 316
 QY 1240 GATGAGCTCTTTCAGAAAACGCGAGCTCAGCAGGAAGCTAAATGAGACCTCGAGATGCT 1299
 DB 317 SerGluGlyLysGlyAspAlaLysLysAspLysLysGlyLysGlySerLys 336
 QY 1300 GATGCCACCGATTGTACAAAGTATAGGAACAATAGAGCAAGTCTAGCGAAGACCAAGG 1359
 DB 337 LysGly-----LysGluSerAlaThrGluSerGluGly 347
 QY 1360 ATAAGGCTAGATTGAGAAAGCTCCAATCCAGTGGCAGAGCAAACTCAAGATCTTC 1419
 DB 348 GluLysGlyAspAlaLysLys---AspAspLysLysGlyLysLys----- 361
 QY 1420 CAGCCTGTGATAGAGCGCTGTGCTCAAAATGTTATTGGCCCGGGTGTCTCACGCTG 1479
 DB 362 -----GlySerLysLys-----GlyLysGluSerAspSerLys 372
 QY 1480 GCGCAGCCGAGCTCGGTGTACTGCGAGTAATGACTGTATCTCTCAACACGCCCGCAGGACA 1539
 DB 373 AlaGluGlyAspLysGlyAspAlaLysLysAsp---AspLysLysAspLysLysGlySer 391
 QY 1540 ATGAGTTTCTAAGCTCAGGTAAAGAACAGACAGAACCCAAAGCCTAAAGAAAGATGAAGTG 1599
 DB 392 LysLysGlyLysGluSerAlaThrGluSerGluGlyLysLysAspSerLysLysAsp 411
 QY 1600 AAGCAGAGAGCCCGAGTCTTCCGAAATGCGGTGCTCAGGAGGATTAAATCTCTTCT 1659
 DB 412 LysAlaGlyLysAspProThrLysAlaGlyLysGlyAsp-----GluSer 428
 QY 1660 GTGCACAGAGACCGAGCTCCAGAAAAAAGAGACCCACAGTGAAGAGCAGTGTGTGTG 1719
 DB 429 LysAspLysLysAspAlaLysLysAspSerLysLysGluLysLysAspGluLysLys 448
 QY 1720 CTTGCGCGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAGCACCGCCGCTGTG 1779
 DB 1779 -----


```

CC CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=2;
CC CC Name=1;
CC CC IsoId=QBBLG0-1; Sequence=Displayed;
CC CC Name=2;
CC CC IsoId=QBBLG0-2; Sequence=VSP_007762;
CC CC Note=Sequence incomplete. No experimental confirmation
CC CC available;
CC CC -!- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.
CC CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; AK033017; BAC28129.1; -
CC CC EMBL; AK038573; BAC30050.1; -
CC CC EMBL; AK045309; BAC32304.1; -
CC CC EMBL; BC011337; AAH11337.1; -
CC CC MGI; 2444148; 6820402020Rik.
CC CC InterPro; IPR002959; Tudor.
CC CC InterPro; IPR007087; Znf_C2H2.
CC CC Pfam; PF00628; PHD; 1.
CC CC Pfam; PF00096; zf-C2H2; 1.
CC CC SMART; SM00249; PHD; 1.
CC CC SMART; SM00333; TUDOR; 2.
CC CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
CC CC PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
CC CC PROSITE; PS01359; ZF_PHD_1; 1.
CC CC PROSITE; PS00016; ZF_PHD_2; FALSE NEG.
CC CC Transcription regulation; Nuclear protein; DNA-binding; Metal-binding;
CC CC Zinc; Zinc-finger; Alternative splicing.
CC CC ZINC_BIND 257 269 A.T HOOK (BY SIMILARITY).
CC CC ZN_FING 455 485 C2H2-TYPE.
CC CC ZN_FING 657 703 PHD-TYPE.
CC CC DOMAIN 162 259 LYS-RICH.
CC CC DOMAIN 526 555 LYS-RICH.
CC CC DOMAIN 543 555 POLY-LYS.
CC CC VARSP LIC 1 28 MKHPNRRGISFVGAQLEARDLNW -> GAARTVLLS
CC CC VGLRRSRSGAVR (in isoform 2).
CC CC /FTIGVSE_007762.
CC CC CONFLICT 817 817 A -> T (IN REF. 2).
CC CC SEQUENCE 1010 AA; 115279 MW; 2BC32811A520342D CRC64;

Alignment Scores:
Pred. No.: 0.00236 Length: 1010
Score: 188.50 Matches: 145
Percent Similarity: 32.78% Conservative: 71
Best Local Similarity: 22.00% Mismatches: 246
Query Match: 4.02% Indels: 198
DB: 1 Gaps: 32

US-09-787-016a-1 (1-2610) x HCS8_MOUSE (1-1010)
QY 299 CACCTAAGCCATCAACCA---CCAGCAAGAGTTCAGAAAACATGGGTTTTCGAA 355
D 351 HisThrValAspThrAsnProLeuProAspLysSerProSerAlaLysAspSerAlaGlu 370
QY 356 GGACCACTA---TCGCCAACGAGGGCGCAGGGACGCGGAGGCTGACCCCACTGGAGC 412
D 371 GlyGlnLeuLysSerProLeuGluAlaGlyGlnValSerSerAlaLeuThr----- 387
QY 413 CGCCACCCCA-----CAGCAGAGTGGGCTGCTCCCTCGCGCGCAGT 456
D 388 CysHisProIleGlyAspGlyLeuGlyAlaAlaAspLeuGluLeuAsnCysLysSerMet 407

```

```

QY 457 GGGAGGCGAGCCCAAGCGCACTGAG-----CGCTGGAGAGCTTCCTG 498
D 408 GlyGluAsnThrMetLysThrGluProValSerProLeuAlaGluValGlnValSer 427
QY 499 ACCATTGCGGCGCGCGCAGGAGGAGCATGCTGCTCTCCCTGGAG-----GATTCT 552
D 428 ThrVal-----GluValProAsnThrLeuLysLysValAspAsp 440
QY 553 GGTGAGCCCGCTCTGCTCCCGGACAGACGCGGACAGACGCTCCGAGGCGGCGTG--- 609
D 441 SerValThrLeuAsnValProAlaValAspLeuAspHisLysPheArgCysLysValLeu 460
QY 609 ----- 609
D 461 AspCysLeuLysPhePheArgLysAlaLysLeuLeuHisTyrHisMetLysTyrPheHis 480
QY 610 -----GAAAGCGCTTCTGAGACGAGGCGGCGGCGGCTGCTTCCAGAGTGTGAAG 663
D 481 GlyMetGluLysSerProGluProGluGluGlyPro-----GlyLysThrHisValGln 498
QY 664 GAACGACGAGCTCTTCTGAAAGGTGAAAGGGGGATGACACGATGACACCTCCGAT 723
D 499 ThrArg-----GlySerAlaValProAspLysThrSer--- 509
QY 724 AGTCACAGCGATGGCTGACCTTGAAAGAGCTTCAGAAATCGCCTTCGAGGAGCGGAA 783
D 510 -----GlnGluSerLeuThrArgLysArgValSerAlaSerProThrAlaLysGlu 527
QY 784 CAGGAGCCCGCTGAG-----AGGCCCTGAAAGGGATCCAGAGTCGCTCGG 831
D 528 LysGluLysThrLysGluLysLysPheLysGluLeuValArgValLysProLysLysLys 547
QY 832 AAGAGGCGCGGAGGAGGTCGCGCGGAGACTGTGGGCTCCGAG---GCCAGTGACACT 888
D 548 LysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 567
QY 889 GTGAG----- 894
D 568 SerGlnGluProSerProLysThrPheAlaValThrArgCysGlySerSerHisLys 587
QY 895 ---GGCGTCTGCGCCAGTAAGCAGGAGCCGAGAGCAGTACAGGGGTGTGTCGAGGCT 951
D 588 ProGlyValHisMetSerProGlnLeuHisGlySerAspAsnGlyAsnHisLysGlyLys 607
QY 952 GGGAAAGATGACAGAGAGAGTAAGTTGAGGGAAGAGCGGCTCAGGACATC-----AAA 1005
D 608 LeuLysThrCysGluGluAspAsnLeuSerGluSerSerSerGluSerPheLeuTrpSer 627
QY 1006 GATGAGGAGCGCTGGAGAC-----TTGGGCGGACCGAGAGCCCTGAATGTGAGGT 1053
D 628 AspGluGluTyrGlyGlnAspValAspValThrThrAsnProAspGluLeuGluGly 647
QY 1054 -----TACGACCCCAACGCGCTGTATTGCAATTCGCCCGCAGCTCACAACACAGG 1104
D 648 AspAspArgTyrAspPheGluValValArgCysLysGluValGlnGluGluAsnAsp 667
QY 1105 TTTATGATTGCTGACCGCTGTCGAAGAATGTTTCATGGCGATTGTGGGCTTCT 1164
D 668 PheMetIleGlnCysGluGluCysGlnCysTrpGlnHisGlyValCysMetGly----- 685
QY 1165 GAGGCTCGAGGAGGCTTTTGAAGGAAT---GGGGAAGACTATATTCGCCCAACTGC 1221
D 686 -----LeuLeuGluGluAsnValProGluLysTyrThrCysTyrValCys 700
QY 1222 ACCATTCTGAGTGCAGGATGAGACTCATTCAGAAACGCGCAGATCAGCAGGAGCTAA 1281
D 701 -----GlnAspProProGly 705
QY 1282 TGGAGACTGGAGATGCTGATGGCCAGT-----TGTACA 1317
D 706 GlnArgProGlyPheLysTyrTrpTyrAspLysGluTrpLeuSerArgGlyHisMetHis 725
QY 1318 AGTATAGGAACAATAGACAGAGAAGTCTAGCGGAAGACCAAGGGATAAAG----- 1365

```

```

Db      726 GlyLeuAlaPheLeuAspGlnAsnTyrSerHisGlnAsnAlaArgLysIleValAlaThr 745
QY      1366 -----GQTAGAATTGAGAAAGCTGCAAAATCCAGTGGCAGAGAACTCAAG 1413
Db      746 HisGlnLeuLeuGlyAspValGlnArgValIleGlnValLeuHisGlyLeuGlnLeuLys 765
QY      1414 ATC-----TTCAGCCTGTGATA 1431
Db      766 MetSerIleLeuGlnSerArgGluHisProAspLeuGlnLeuTyrCysGlnProTrrLys 785
QY      1432 GAGCGCGCTGGTGGCTCAAAATGATTGGCCCGGGTGTGTACCTGGCGGAGCCGAC 1491
Db      786 GlnHisSerGlyGluGlyArg-----AlaHisProArg 796
QY      1492 TCGGTGACTGCTAGTAAGTACTGTATCTCTCAACACGCCCGCAGCAGATGAAGTTCTTA 1551
Db      797 HisIleHis-----Ile 800
QY      1552 AGCTCAGGTAAAGACAGACGCAAGCCCTAAAGAAAGATGAAGTGAAGCCAGAGAAAG 1611
Db      801 ThrAspAlaArgSerGluGluGluSerProSerTyrArgThrLeuAsnGlyAlaValGluLys 820
QY      1612 CCCAGT---CTTCGGAATCGGTGCTCAGCGCAGGTATTAAATCTCTTCTGTGCACAAG 1668
Db      821 ProSerProLeuProArgSerValGluGluSer-----TyrIleThrSerGluHisCys 838
QY      1669 AGACCAGCTCCAGAAAAAGACACACACATGAGAGGACGTCGTGCTCCCTGCGCGG 1728
Db      839 TyrGlnLysProArgAlaTyrTyrProAlaValGluGlnArgLeuValValGluThrArg 858
QY      1729 AGTGAACACTCCGGAAGGAGCAGCT-----TGTGAGAGCAGCAGCGCGTGC----- 1776
Db      859 GlySerAlaLeuAspAlaAlaValSerProLeuGluGluGluGluGluGluGluGluGlu 878
QY      1777 -----TGGCGCAGCATCAACAATTACATCA-----GTAAAGCCCA 1812
Db      879 ProArgLeuGlyTyrProIleAspGlnAspArgSerArgGlyAspIleAspProLysPro 898
QY      1813 -----GAAAGACTGCTGCTCCTCCCTCGCGTCACTGCTTG 1845
Db      899 SerSerProLysValArgGluTyrIleSerLysAsnValLeuProGluGluThrProAla 918
QY      1846 TATAAATGATGATNCACCTAGGGTGGCTCTCTGAGCCCTCCCTCTTCTTCTG 1902
Db      919 ArgIleLeuLeuAspArgGlyGlyGluGluValSer---SerGlnHisGlnTrrp 936

RESULT 9
NP14_RAT
ID      NP14_RAT          STANDARD;          PRT;          704 AA.
AC      P41777;
DT      01-NOV-1995 (Rel. 32, Created)
DT      01-NOV-1995 (Rel. 32, Last sequence update)
DT      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Nucleolar phosphoprotein p130 (Nucleolar 130 kDa protein) (140 kDa
DE      nucleolar phosphoprotein) (Nopp140) (Nucleolar and coiled-body
DE      phosphoprotein 1).
DE      NCBI.
GN      Rattus norvegicus (Rat).
OS      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 292-309 AND 563-601.
RC      TISSUE=Liver;
RX      MEDLINE=92323542; PubMed=1623516;
RA      Meier U.T., Blobel G.;
RT      "Nopp140 shuttles on tracks between nucleolus and cytoplasm.";
RL      Cell 70:127-138(1992).
RN      [2]
RP      INTERACTION WITH NOPS AND FIBRILLARIN.
RX      MEDLINE=20143579; PubMed=10679015;
RA      Yang Y., Isaac C., Wang C., Dragon F., Pogacic V., Meier U.T.;

```

```

RT      "Conserved composition of mammalian box H/ACA and box C/D small
RT      nucleolar ribonucleoprotein particles and their interaction with the
RT      common factor Nopp140.";
RL      Mol. Biol. Cell 11:567-577(2000).
CC      -!- FUNCTION: Related to nucleogenesis, may play a role in the
CC      maintenance of the fundamental structure of the fibrillar center
CC      and dense fibrillar component in the nucleolus. It has intrinsic
CC      GTPase and ATPase activities. May play an important role in
CC      transcription catalyzed by RNA polymerase I (By similarity).
CC      -!- SUBUNIT: Interacts with DKC1/NAP57, NOP5/NAP65 and fibrillarin.
CC      -!- SUBCELLULAR LOCATION: Shuttles on curvilinear tracks between
CC      nucleolus and cytoplasm. These tracks extend from the dense
CC      fibrillar component of the nucleolus across the nucleoplasm to
CC      a limited number of nuclear pore complexes.
CC      -!- PTM: Undergoes rapid and massive phosphorylation/dephosphorylation
CC      cycles on CK2 and PKC sites. Nopp140 is one of the mostly
CC      phosphorylated proteins in the cell.
CC      -!- SIMILARITY: Contains 1 Lish domain.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; M94287; AAA41718.1; --
CC      EMBL; M94288; AAA41719.1; --
CC      PIR; B42680; B42680; Lish.
CC      InterPro; IPR006594; Lish.
CC      InterPro; IPR007718; SRP40.C.
CC      InterPro; IPR003993; treacle.
CC      Pfam; PF05022; SRP40.C; 1.
CC      PRINTS; PR01503; TREACLE.
CC      SMART; SM00667; Lish; 1.
CC      PROSITE; PS00836; Lish; 1.
CC      Nuclear protein; Phosphorylation; Repeat; GTP-binding; ATP-binding.
KW      DOMAIN 10 42
FT      DOMAIN 84 570
FT      REPEAT 84 95
FT      REPEAT 127 138
FT      REPEAT 170 181
FT      REPEAT 231 242
FT      REPEAT 274 285
FT      REPEAT 335 346
FT      REPEAT 373 384
FT      REPEAT 434 445
FT      REPEAT 479 490
FT      REPEAT 524 535
FT      REPEAT 559 570
FT      MOD RES 567 567
FT      VARIANT 150 150 MISSING (IN NOPP140B).
SQ      SEQUENCE 704 AA; 73562 MW; 14DF1BF2DE483EA3 CRC64;

Alignment Scores:
Pred. No.: 0.00466 Length: 704
Score: 182.50 Matches: 134
Percent Similarity: 35.42% Conservative: 75
Best Local Similarity: 22.71% Mismatches: 251
Query Match: 3.89% Indels: 130
DB: Gaps: 25

US-09-787-016A-1 (1-2610) x NP14_RAT (1-704)
QY      172 TCGCTTGGACCCAGAGGTTTGGTGGCAGGGTTTTGTTGTTATTAGGATTCAGGG 231
Db      55 SerPheTrpLeuLysSerThrLysAlaPro-----LysValLysLeuGlnSerArgGly 72
QY      232 AAAAGTGTCCACGCTTTCAGTGTGTGGAGCAGGTATGGACGACAAAGCCGCGAGCAAT 291
Db      73 ProValAlaLysLysAlaLysLysLysGluThrSerSerSerSerSerSerSerSerSer 92

```

[illegible][illegible]

FT	REPEAT	84	95	ACIDIC SERINE CLUSTER 1.	
FT	REPEAT	125	136	ACIDIC SERINE CLUSTER 2.	
FT	REPEAT	167	178	ACIDIC SERINE CLUSTER 3.	
FT	REPEAT	221	232	ACIDIC SERINE CLUSTER 4.	
FT	REPEAT	264	275	ACIDIC SERINE CLUSTER 5.	
FT	REPEAT	325	336	ACIDIC SERINE CLUSTER 6.	
FT	REPEAT	363	374	ACIDIC SERINE CLUSTER 7.	
FT	REPEAT	425	436	ACIDIC SERINE CLUSTER 8.	
FT	REPEAT	470	481	ACIDIC SERINE CLUSTER 9.	
FT	REPEAT	519	529	ACIDIC SERINE CLUSTER 10.	
FT	REPEAT	555	566	ACIDIC SERINE CLUSTER 11.	
FT	DOMAIN	68	82	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).	
FT	DOMAIN	204	382	INTERACTS WITH RPA194.	
FT	DOMAIN	384	587	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).	
FT	DOMAIN	601	617	NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).	
FT	MOD_RES	563	563	PHOSPHORYLATION (BY CK2) (BY SIMILARITY).	
FT	VARSPLIC	241	241	K -> KWTITVSRAE (in isoform Beta).	
FT				/FTIG=VSP_004338.	
FT	CONFLICT	3	3	D -> A (IN REF. 2).	
FT	CONFLICT	133	133	R -> S (IN REF. 2).	
FT	CONFLICT	291	292	YA -> SV (IN REF. 2).	
FT	CONFLICT	456	456	S -> P (IN REF. 2).	
FT	SEQUENCE	699 AA;	73720 MW;	DFD4AD94EDF659PB CRC64;	
DB:					
Alignment Scores:					
Pred. No.:		0.00437	Length:	699	
Score:		182.00	Matches:	117	
Percent Similarity:		36.10%	Conservative:	83	
Best Local Similarity:		21.12%	Mismatches:	213	
Query Match:		3.88%	Indels:	141	
		1	Gaps:	19	
US-09-787-016A-1 (1-2610) x NP14_HUMAN (1-699)					
QY	268	GACGACAAAGGCGACCCGACGCAATGAGGAGCGCACCTAAGCGCCATCAACCCACACGACAA	327		
DB	136	ASpAspGluaspGlnLysGlnProValGlnLysGlyValLysProGlnAlaLys	155		
QY	328	GAGTTCAGGAAACATGGGTTTCGAGAGCACCTATCGCCAGCAGGAGGCGCGCAGG	387		
DB	156	AlaAlaLysAlaPro-----	167		
QY	388	GAGCGGAGGCTGAC---CCACTGGAGCGCGCCACCCACACGACGACGAGCTGGCGCTGCC	444		
DB	168	AspSerAspSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	183		
QY	445	CTGGCGGCGCAGTGGGAGGCGACCCGACGCGACTGAGCGCGTGGAGCAGCTTCTGACCAT	504		
DB	184	-----LysProLysIleThr-----	188		
QY	505	GCGCGGCGCGCGGCGGAGGAGCATGCTGCTCTCCCTGGAGGATTCGTGGAGCCACG	564		
DB	189	-----ProValThrValLysAlaGlnThrLysAlaPro	199		
QY	565	TCCTGCCCCGCGACAGACGCGGAGCAGCTCCGAGGCGCAGCTGCGAAAGCGCTTCTGAG	624		
DB	200	ProLysProAlaAlaAlaProLysIleAlaAsnGlyLysAlaAlaSerSerSerSer	219		
QY	625	ACCAGAGCGGCCCGAGCTGCTCCACAGCTGTGAGGAGAACGACGACCTCTTCTGAA	684		
DB	220	Ser	239		
QY	685	AAG-----	705		
DB	240	LysLysThrValProLysLysGlnValValAlaLysAlaProValLysAlaAlaThrThr	259		
QY	706	CACGATGACCTCGGATGACGCGATGGCTGACCTTGAAGAGCTTCAGAAATCGC	765		
DB	260	ProThrArgLysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer	279		
QY	766	CTTGCAGGAAGCGGGAACAGGAGCC-----	801		
DB	280	LysProMetLysAsnLysProGlyProLysSerTyAlaProProProSerAlaProPro	299		

```
QY 802 CCCCTGAAGGATCCAGAGTCCTGCGGAAGAAGCGCGGAGGAGGTCCTCCGCCGAG 861
D 803 ||||| : : : : : ||||| : : : : :
D 300 ProLysLysSerLeuGlyThrGlnProProLysLysAlaValGluLysGlnProVal 319
QY 862 ACTGTGGCTCCAGGCGAGTACACTGTGGAGGGCTCTCCAGTAAAGCAGAGGCC 921
D 863 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 320 GluSerSerGluAspSerSerAspGluSerSerSerSerGluGluGluLysPro 339
QY 922 GAGAACGATCAGCGGGTGTGTCCAGGCT : : : : : : : : : : : : : : : : : : : : :
D 923 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 340 ProThr---LysAlaValSerLysAlaThrThrLysProProAlaLysLysAla 358
QY 964 AGAGAGAGTAAGTTCGAGGAAGCGGCTCAGGACATCAAGATGAGAGAG 1014
D 965 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 359 AlaGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 378
QY 1015 -----CCTGGAGACTGGCGCCAGCGAAGCTGAATGTGAGGCTAGCACCCCAAGCC 1068
D 1016 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 379 LysProAlaGly-ThrThrLysAsnSerSerSerSerSerProAlaValThrThrLysSerPr 398
QY 1069 CTGTATTGCTTGGCGAGCTCACAACAGAGTTATGATTGTGTGACCGCTGT 1128
D 1070 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 398 AlaValLysProAlaAla-AlaProLysGlnProValGly----- 411
QY 1129 GAAGATGCTTTCATGCGGATGTGTGGCAATTTCTGAGGCTCAGGAGGCTTTTGGAA 1188
D 1130 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 412 -----GlyGlnGln-LysLeuLeuThr 418
QY 1189 AGAATGGGGAAGACTATATCTGCCAACTGCACCACTTCTCAAGTGCAGATGACT 1248
D 1190 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 419 ArgLysAlaAsp-----SerSerSerSerGluGlu 428
QY 1249 CATTGAGAAACCGCAGATCAGCAGGAGCTAAA-----TGGAGACTGGA 1293
D 1250 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 429 GluSerSerSerSerGluGluLysThrLysLysMetValAlaThrThrLysPro--- 447
QY 1294 GATGTGATGGCAGCGATTTGTAAGATATA-----GGAAACAATA 1332
D 1295 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 448 LysAlaThrAlaLysAlaAlaLeuSerLeuProAlaLysGlnAlaProGlnGlySerArg 467
QY 1333 GAGCAGAGTCTAGCAGACAGCAGGATAAAGGTAGATTGAGAAGACTCCAAATCCA 1392
D 1334 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 468 AspSerSerSerSerSerSerSerSerSerSerSerGluGluGluGluLysThrSerLys 487
QY 1393 AGTGGCAAGCAAACTCAAGATCTTCCAGCTGTGTATAGAGCGCTGGTCCCTCAAAA 1452
D 1394 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 488 SerAlaValLysLys-----LysProGlnLysValAlaGlyGlyAlaAlaPro 503
QY 1453 TGTATTGGCCCGGTGTGTACGTGGCGGCGGAGCTCGGTGTACTGCGAGTAATGAC 1512
D 1454 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 504 -----SerLysProAlaSerAla----- 509
QY 1513 TGTATCTCAACACGCGCAGCACAATGAAGTTTCTAAGCTCAGGTAAAGACAGAG 1572
D 1514 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 510 -----LysLysGlyLysAlaGluSerSerSerSerSerSerSerSerSerSerSer 526
QY 1573 CCAGAGCTTAAAGAAAGATGAAGTGAAG-----CCAGAGAACCCAGTCTTCCGAATGC 1629
D 1574 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 527 GluGluGluGluLysLysLysGlyLysGlySerProArgProGlnAlaProLysAla 546
QY 1630 GTGTCTCAGCAGGATTAATAATCTCTCTGTGTGCACAGAGCAGCTCCAGAAAAA 1689
D 1631 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 547 AsnGlyThrSerSerAlaThrAlaGlnAsnGlyLysAlaAlaLysAsnSerGluGlu 566
QY 1690 GAGACACAGTGAAGAGGAGTGTGTGCTCCCTGGCGGAGTGAAGACTCGGGAG--- 1746
D 1691 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 567 GluGluGluLysLysLysAlaAlaValValSerLysSerGlySerLeuLysLysArg 586
QY 1747 -----GAGCAGCTTGTGTGAGCAGCAGCGCG 1773
D 1748 ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 587 LysGlnAsnGluAlaLysLysGluAlaGluThrPro 598
```

```
RESULT 11
MYS3_HUMAN
ID MYS3_HUMAN STANDARD; PRT; 2004 AA.
AC Q92794;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE MYST histone acetyltransferase 3 (Runt-related transcription factor
DE binding protein 2) (Monocytic leukemia zinc finger protein) (Zinc
DE finger protein 220).
GN MYST3 OR RUNXBP2 OR ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=96376968; PubMed=9782817;
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
RA Horman D., Mitalman P., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -!- FUNCTION: May represent a chromatin-associated acetyltransferase.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DISEASE: Involved in acute myeloid leukemias through a chromosomal
CC translocation t(8;16)(p11;p13) involving MYST3 and CREBBP.
CC -!- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC -!- SIMILARITY: Belongs to the MYST (SAS/MOZ) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC SML; U47742; AAC50662.1; --
CC Genew: HGNC:13013; MYST3.
CC
CC MIM; 601408; --
CC GO; GO:0006323; P: DNA packaging; TAS.
CC InterPro; IPR005818; Histone_H1/H5.
CC InterPro; IPR003717; MOZ_SAS.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF01853; MOZ_SAS; 1.
CC Pfam; PF00628; PHD_2.
CC SMART; SM00526; H15; 1.
CC SMART; SM00249; PHD; 2.
CC PROSITE; PS01359; ZF_PHD_1; 1.
CC PROSITE; PS0016; ZF_PHD_2; 2.
CC Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW Nuclear protein.
FT ZN_FING 206 265 PHD-TYPE 1.
FT ZN_FING 259 313 PHD-TYPE 2.
FT DOMAIN 371 379 POLY-SER.
FT ZN_FING 538 560 C2HC-TYPE.
FT DOMAIN 788 801 POLY-GLU.
FT DOMAIN 989 995 POLY-GLU.
FT DOMAIN 1019 1026 POLY-ARG.
FT DOMAIN 1069 1078 POLY-GLU.
FT DOMAIN 1147 1150 POLY-LYS.
FT DOMAIN 1221 1242 GLU-RICH.
FT DOMAIN 1267 1302 POLY-GLU.
FT DOMAIN 1411 1414 POLY-GLU.
FT DOMAIN 1593 1597 POLY-SER.
FT DOMAIN 1643 1704 GLN/PRO-RICH.
FT DOMAIN 1897 1977 MET-RICH.
FT SITE 1546 1547 BREAKPOINT FOR TRANSLOCATION TO FORM
FT MOZ-CBB.
FT SEQUENCE 2004 AA; 225054 MW; 9FEBBAC3792854BA CRC64;
```


QY 1527 C---GCCGAGCGACATGAGTTCTAGCTCAGTAAAGAACGAGACGCTTAA 1583
 Db 1634 rProGlnSerCysValGluGlyProSerAsnGlnGlnGlnProProProPr 1654
 QY 1584 AGAAAAGATGAAGTGAAGCCAGAGAGCCAGTCTTCGGAATCGGTGCTCAGGCAGG 1643
 Db 1654 oProProGlnGlnProGlnProProProGln---ProGlnProAlaProGlnProPr 1673
 QY 1644 TATTAAATCTCTTCTGTCACAGAGAGACGCTCAGAAAAAAGAG 1692
 Db 1673 oProProGlnGlnGlnProGlnGlnGlnProGlnProGlnProGlnGln 1689
 RESULT 12
 HC58 HUMAN
 ID Q9BWT0; Q9BWA4; Q9BXA3; Q9BZW3; Q9H421; Q9H4J6; Q9WZ22;
 AC 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Hepatocellular carcinoma-associated antigen 58 (Glioma-expressed
 antigen 2) (Transcription factor T2P) (Novel zinc finger protein).
 GN HCA58 OR GLEA2 OR T2P OR NZF OR C20ORF104.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2), TISSUE SPECIFICITY, AND DISEASE.
 RC TISSUE=Hepatoma;
 RX MEDLINE=22092308; PubMed=12097419;
 RA Wang Y., Han K.-J., Pang X.-W., Vaughan H.A., Qu W., Dong X.-Y.,
 RA Peng J.-R., Zhao H.-T., Rui J.-A., Leng X.-S., Cebon J., Burgess A.W.,
 RA Chen W.-Z.;
 RT "Large scale identification of human hepatocellular carcinoma-
 associated antigens by autoantibodies.";
 RL J. Immunol. 169:1102-1109(2002).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RA Cheng J.Q., Kaneko S., Dan H.C., Testa J.R.;
 RT "Cloning and characterization of a novel transcription factor
 (T2P)."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 RA Buck D., Burrill W.D., Butler A.P., Gardner C., Carter N.P.,
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.B.,
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle B., Hunt A.R., Hunt S.E., Jekesch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Levanthalo M.H., Leverhulme M., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.L., Martin S.L., McConachie I.J., McLay K., McMurray A.A.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [4]

RP SEQUENCE OF 1-550 FROM N.A. (ISOFORM 1).
 RC TISSUE=Eye, and Lymph;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Staubsberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Kleinsch R.D., Collins F.S., Wagner L., Schenken C.F., Bhat N.K.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Max S.I., Wang J.M., Hsieh F.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.M., Hong L.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uedini T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raka S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Paray J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Maria M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [5]
 RP SEQUENCE OF 1-543 FROM N.A. (ISOFORM 1), AND DISEASE.
 RC TISSUE=Glioblastoma;
 RX MEDLINE=21560358; PubMed=11703362;
 RA Fischer U., Struss A.-K., Hemmer D., Pallasch C.P., Steudel W.-I.,
 RA Meese E.;
 RT "Glioma-expressed antigen 2 (GLEA2): a novel protein that can elicit
 immune responses in glioblastoma patients and some controls.";
 RL Clin. Exp. Immunol. 126:206-213(2001).
 RN [6]
 RP DISEASE.
 RX MEDLINE=22683448; PubMed=12800201;
 RA Behrends U., Schneider I., Roessler S., Frauenknecht H., Golbeck A.,
 RA Lechner B., Eigenstetter G., Zobywalski C., Mueller-Weirich S.,
 RA Graubner U., Schmid I., Sackner J., Spaeth M., Goetz C., Prantl P.,
 RA Assmus H.-P., Bise K., Mautner J.;
 RT "Novel tumor antigens identified by autologous antibody screening of
 childhood medulloblastoma cDNA libraries.";
 RL Int. J. Cancer 106:244-251(2003).
 CC - FUNCTION: Possible transcription factor.
 CC - SUBCELLULAR LOCATION: Nuclear (Probable).
 CC - ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9BVI0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9BVI0-2; Sequence=VSP 007760, VSP 007761;
 CC Note=Ref.1 isoform 2 is in conflict in position(s):
 CC 237-239:KTR-SKTS;
 CC - TISSUE SPECIFICITY: Expressed in heart, kidney, liver, lung,
 CC pancreas, placenta, spleen and testis. Not expressed in brain,
 CC skeletal muscle, colon, ovary, prostate, small intestine and
 CC thymus. Expressed in colon and ovary cancer cell line, whereas it
 CC is not expressed in the respective normal tissues.
 CC - DISEASE: Antibodies against C20orf104 are present in sera from
 CC patients with hepatocellular carcinoma, glioblastoma and childhood
 CC medulloblastoma who developed different autoantibodies. May serve
 CC as diagnostic marker in glioblastoma.
 CC - SIMILARITY: Contains 1 A/T hook DNA-binding repeat.
 CC - SIMILARITY: Contains 1 PHD-type zinc finger.
 CC - SIMILARITY: Contains 1 C2H2-type zinc finger.
 CC - CAUTION: Ref.1 sequence differs from that shown due to a
 CC frameshift in position 260.
 CC - CAUTION: Ref.2 (AAK19748) sequence differs from that shown due to
 CC frameshifts in positions 635, 653 and 655.
 CC - CAUTION: Ref.4 sequence differs from that shown due to a chimeric
 CC cDNA.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@ibp-sib.ch).

ENBL;	AF220416;	AAP34184.1;	ALT_FRAME.
DR	ENBL;	AY027523;	AAK13046.1; -
DR	ENBL;	AF348207;	AAK19748.1;
DR	ENBL;	AL078461;	CAC09384.1; -
DR	ENBL;	AL109965;	CAC09389.1; -
DR	ENBL;	AF258787;	AAG49888.1; -
DR	ENBL;	BC001198;	AAXH01198.1;
DR	ENBL;	BC048210;	AAX48210.1; -
DR	Gnew;	HGNC:16098;	C20orf104.
DR	InterPro;	IPR002999;	Tudor.
DR	InterPro;	IPR007087;	Znf_C2H2.
DR	InterPro;	IPR001965;	Znf_PHD.
DR	Pfam;	PF00628;	PHD; 1.
DR	Pfam;	PF00096;	zf-C2H2; 1.
DR	SMART;	SMO0249;	PHD; 1.
DR	SMART;	SMO0333;	TUDOR; 2.
DR	PROSITE;	PS00028;	ZINC_FINGER_C2H2_1; 1.
DR	PROSITE;	PS0157;	ZINC_FINGER_C2H2_2; 1.
DR	PROSITE;	PS01359;	ZF_PHD_1; 1.
DR	PROSITE;	PS50016;	ZF_PHD_2; FALSE NEG.
XW	Transcription regulation; Nuclear protein; DNA-binding; Metal-binding;		
KW	Zinc; Zinc-finger;	Alternative splicing.	
FT	DNA_BIND	257	269 A.T.HOOK
FT	ZN_FING	452	477 C2H2-TYPE.
FT	ZN_FING	654	700 PHD-TYPE.
FT	DOMAIN	162	221 LYS-RICH.
FT	DOMAIN	523	552 LYS-RICH.
FT	VARSPLIC	237	273 VDKPENDINGVKSPQENLRPKRKGRPPSTAPTAVDS -> KTROTPTFHSSVCCGFKLNSFNATNIGTEKKENIKRM (in isoform 2).
FT		/ftid=vsp_007760.	
FT	VARSPLIC	274	1012 Missing (in isoform 2).
FT		/ftid=vsp_007761.	
FT	CONFLICT	221	222 KE -> QG (IN REF. 5).
FT	CONFLICT	226	226 E -> G (IN REF. 5).
FT	CONFLICT	516	516 S -> F (IN REF. 5).
SO	SEQUENCE	1012 AA:	1115385 MW: 1CDBADC23D007503 CRC64;

UIS-09-787-016A-1 (1-2610) x HC58 HUMAN (1-1012)

Qy	271	GACAAAGCGGACCCGAGCAATGAGAGGCACCTTAAGGCCATCAACCCACGACGAAGAG	330
		:::	:::
Db	204	GluLysGlyLysValSerGluSerLeuProLysAsnGluLysGluAspLysGluAsn	223
		:::	:::
Qy	331	TTCAGGAAACATCGGGTTTTCGAAGGACCACTATCGCAAGCGAGAGGGCGCAGGGGAC	390
		:::	:::
Db	224	IleSerGluAsn-----AspArgGluTyrSerGlyAsp	234
		:::	:::
Qy	391	CGCGAGGCTGACCCACTGGAGCG-----CCACCCCCACGACGACAGCTG	435
		:::	:::
Db	235	AlaGlnValAspLys---LysProGluAsnAspIleValLysSerProGlnGluAsnLeu	253
		:::	:::
Qy	436	GGCCTGTCTCTCGCGCGCAGTGGAGGAGCGCCCAAG-----CGC	474
		:::	:::
Db	254	ArgGluProLysArgLysArgGlyArgProProSerIleAlaProThrAlaValAspSer	273
		:::	:::
Qy	475	ACTGAGCGCTGGAGCAGTCTCTCCATCCATTGCGCGCGCGCGC-----	516
		:::	:::
Db	274	AsnSerGlnThrLeuGlnProIleThrLeuGluLeuArgArgLysIleSerLysGly	293
		:::	:::

QY	1066	GCCTGTATTCAGTTCGCGCCAGCGCTCACAAACAGGTTTATGATTCGTGACGCGC	1125
DB	652	ValValArgCysIleCysGluValGlnGluAsnAspPheMetIleGlnCysGluGlu	671
QY	1126	TGTGAAGATGTTTCATGCGCATGTGTGGCATTCTGAGCTCGAGGAGGCTTTTG	1185
DB	672	CysGlnCysTrpGlnHisGlyValCysMetGly	684
QY	1186	GAAAGGAAT---GGGGAAGACTATATCTGCCAAACTGCACCATCTCTCAAGTCAGCAT	1242
DB	685	GluGluAsnValProGluLysTrpCysTrpValCys	697
QY	1243	GAGACTCATTCAGAAACGCGCAGATCAGCAGGAGCTAAATGAGACCTGGAGATCTCAT	1302
DB	698	-----GlnAspProProGlyGlnArgProGlyPheLysTrp	709
QY	1303	GGCACCAGT-----TGACAAAGTATAGCAACAATACAGCAG	1338
DB	710	TrpTrpAspLysGluTrpLeuSerArgGlyHisMetHisGlyLeuAlaPheLeuGluGlu	729
QY	1339	AAGTCTAGCGAAGCAACAGGGATAAG-----GGTGAAT	1374
DB	730	AsnTyrSerHisGlnAsnAlaLysLysIleValAlaThrHisGlnLeuLeuGlyAspVal	749
QY	1375	GAGAAAGCTGCAAT-----CCAAGTGGCAAGAACTCAAGATCTTCAGCCTGTG	1428
DB	750	GlnArgValIleGluValLeuHisGlyLeuGlnLeuLysMetSerIleLeuGlnSer	768
QY	1429	ATAGAGGCGCTGTGCTCCCTCAAAATGTATGTGCCCGGGTGTGTACGTGCGCGACCC	1488
DB	769	ArgGluHisProAspLeuProLeuTyrCysGlnProTyrLysGlnHisSerGlyGluGly	788
QY	1489	GACTGGGTACTGCACT-----ANTGACTGTATCTCAACACGCCCGCAGCACA	1539
DB	789	ArgSerHisPheArgAsnIleProValThrAspThrArgSerLysGluAlaProSer	808
QY	1540	ATGAGTCTTCAAGCTCAGTTAAGCAACAGACCAAGCCAAAGCTTAAAGAAAGATG	1599
DB	809	TyrArgThrLeu---AsnGlyAlaValGluLysProArgProLeu	822
QY	1600	AAGCCAGAGACCCAGCTTCGGAATCGGTGCTCAGCGAGGTATTAATCTCTCTCT	1659
DB	823	-----AlaLeuProLeuProArgSerValGluGluSerTyrIleThrSer	837
QY	1660	GTGCACAGACACAGCTCCAGAAAAAAGAGACCAAGTGAAGACGAGTGTGTGTC	1719
DB	838	GluHisCysTyrGlnLysProArgAlaTyrTyrProAlaValGluGlnLysLeuValVal	857
QY	1720	CTGCGCGGAGTGAACATCTCGG-----AAGAACAGCTTGTGTGAG	1761
DB	858	GluThrArgGlySerAlaLeuAspAspAlaValAsnProLeuHisGluAsnGlyAspAsp	877
QY	1762	AGCAGCACCG-----TCGTGGCGGAGCATCAATTAATGCA-----	1803
DB	878	SerLeuSerProArgLeuGlyTrpProLeuAspGlnAspArgSerLysGlyAspSerAsp	897
QY	1804	GTAAGCCA-----GAAAGACTCTCTCTCCCTCGCGC	1836
DB	898	ProLysProGlySerProLysValLysGluTyrValSerLysLysAlaLeuProGluGlu	917
QY	1837	TCAGTGTGTATAAAGTATGATACCTAGGGGTGGCTCTGACCCCTCCGCTCT	1896
DB	918	AlaProAlaArgLysLeuLeuAspArgGlyGlyGluGlyLeuLeuSer---SerGlnHis	936
QY	1897	TTCTGG	1902
DB	937	GlnTrp	938
RESULT 13			
ID	TRDN	CANFA	
AC	P82179	STANDARD;	PET; 700 AA.
DT	28-FEB-2003	(Rel. 41, Created)	

DT	28-FEB-2003	(Rel. 41, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DB	Triadin.		
GN	TRDN		
OS	Canis familiaris (Dog)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
OX	[1]		
RN	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.		
RP	TISSUE=Heart, and Skeletal muscle;		
RC	MEDLINE=99426545; PubMed=10497235;		
RA	Kobayashi Y.M.; Jones L.R.;		
RT	"Identification of triadin 1 as the predominant triadin isoform		
RT	expressed in mammalian myocardium."		
RL	J. Biol. Chem. 274:28660-28668(1999).		
CC	!- FUNCTION: May be involved in anchoring calsequestrin to the		
CC	junctional sarcoplasmic reticulum and allowing its functional		
CC	coupling with the ryanodine receptor (By similarity).		
CC	!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic		
CC	reticulum.		
CC	!- ALTERNATIVE PRODUCTS:		
CC	Event=Alternative splicing; Named isoforms=3;		
CC	Name=Skeletal;		
CC	Isoid=P82179-1; Sequence=Displayed;		
CC	Isoid=P82179-2; Sequence=VSP_004001, VSP_004002;		
CC	Name=Cardiac 3;		
CC	Isoid=P82179-3; Sequence=VSP_004003, VSP_004004;		
CC	!- TISSUE SPECIFICITY: Skeletal and cardiac muscle.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL; AF165916; AAF00222.1; --		
DR	EMBL; AF165915; AAF00221.1; --		
DR	EMBL; AF165917; AAF00223.1; --		
KW	Transmembrane; Sarcoplasmic reticulum; Glycoprotein;		
KW	Alternative splicing.		
FT	INIT MET	0	
FT	DOMAIN	1 46	BY SIMILARITY.
FT	TRANSMEM	47 67	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	68 700	POTENTIAL.
FT	CARBOHYD	74 74	LUMENAL (POTENTIAL).
FT	CARBOHYD	616 616	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	VARSPLIC	257 277	DOYAFCRYWIDMFVHGDLPFG -> GKHSSEVAGGSKRTLG
FT			KKQIQ (in isoform Cardiac 1).
FT			/FTid=VSP_004001.
FT			Missing (in isoform Cardiac 1).
FT	VARSPLIC	278 700	/FTid=VSP_004002.
FT			E -> EPIKGKVKVPGSLKEKE (in isoform
FT			Cardiac 3).
FT			/FTid=VSP_004003.
FT			EKKVVKQVKAIEKAAIEKTVKPKAKAEHQEKESPTIKTD
FT			KPKTSKETPSVETES -> GILQVVPVINCFLVQFOODE
FT			ELNVESKVRMIHVLHSPTSRSTSPILVISTCTKT (in
FT			isoform Cardiac 3).
FT			/FTid=VSP_004004.
FT			
SQ	SEQUENCE	700 AA; 78152 MW; P033E3A1BEE0C56 CRC64;	
Alignment Scores:			
Score:	0.0139	Length:	700
Pred. No.:	174.00	Matches:	107
Percent Similarity:	38.49%	Conservative:	77
Best Local Similarity:	22.38%	Mismatches:	190
Query Match:	3.71%	Indels:	104
DB:	1	Gaps:	20

Qy	604	AGCGTGAAGACGGCTTCTGACACAGACGCGCCCC---CAGTCTGCTTCCACAGCTGTG	660
Db	76	SerAlaSerSerLeuSerLysIleGlySerAspProLeuLysLeuValHisAspAlaVal	95
Qy	661	AAGGAACGACCCAGCGCTCTTCTGAAGAAAGGTGAAGA	696
Db	96	GluGlu-----ThrThrAspTrpValIyrGlyPhePheSerLeuSerAspIle	112
Qy	697	-----GGGGATGACCAAGATGACACCTCCGATAGTACAGCGATGGCCTGACC	744
Db	113	IleSerSerAspGlyAspGluAspAspAspAspGlyAspGluAspThrAspGlyGlu	132
Qy	745	TTGAAGAG-----CTTCAGAAATCGCTTCGCGAGGACGGGACACAGGACCCCACTGAG	798
Db	133	IleGluGluProProLeuLysGlnLysGluIleHisLysGluLysAlaGluLysGluGlu	152
Qy	799	AGGCGCCTGAAGGGATCCAGAGTCGCTCGGGAAGAGCGCCGGAGGAGGCTCCCGCC	858
Db	153	LysProGluArgLysIleLeuAlaLysValAlaHisArgGluLysGluLys---ValLys	177
Qy	859	GAGACTGTGGGCTCCGAGGCGAGTGACACTGTGGAGGGCGTCCTGCCCACTAAGCAGGAG	918
Db	172	GluLysGluLysSerGluLysLysAlaThrHisLysGluLysIleGluLysLysGluLys	191
Qy	919	CCCGAAGACGATCAGGGGTTGTGCCAGGCTCGGAAGATGACAGAGAGTAAGTTG	978
Db	192	ProGlu-----ThrLysThrMetAlaLysGluGluArgLysAlaLysThr	206
Qy	979	GAGGGAAGCGGCTCAGGACATCAAGATGAGGAGCGCTGAGACTTGGCGCCAGCGGAAG	1038
Db	207	GluGluLysIleLysLysGluValLysGly-----GlyLysGlnLysValLys	223
Qy	1039	CCT-----GAATGTGAG	1050
Db	224	ProThrAlaAlaLysValLysGluValGlnLysThrProProLysAlaLysGluLysGlu	243
Qy	1051	GGTTACGACCCCAACGCGCTGTATTGCAATTTGCCCGCCAGCTCACAACACAGGTTTATG	1110
Db	244	GlyLysGluThrAlaAla-----ValAlaLysHisGluGlnLysAspGlnTyrAla	260
Qy	1111	ATTTCCTGTGACCGCTGGAAGATGGTTTCATGCCGATTCT---GTGGCATTTCTGAG	1167
Db	261	PheCysArgTyrMetIleAspMetPheValHisGlyAspLeuArgProGlyLysSerPro	280
Qy	1168	GCTCGAGGGAGCGCTTTTGGAAAGAAATGGGAAGACTATATCTGCCCAACTGCACCANT	1227
Db	281	AlaLeuProProProLeuProThrValGlnAlaSerArgProThrProAlaSerProThr	300
Qy	1228	CTGCAAGTGCAGGATGAGACTCATTCCAGAAACGGCAGATCAGCAGGAAGCT-----	1278
Db	301	LeuGluGlyLysGluGluGluGluLysLysLysAlaGluLysValThrSerGluThr	320
Qy	1278	-----1278	-----
Db	321	LysLysLysGluLysGluAspValLysLysLysSerAspLysAspThrAlaIleAspVal	340
Qy	1279	---NAATGGAGACCTCGAGATGCTCATGGCCACCGATTGTACAAGTATAGGACATA---	1332
Db	341	GluLysLysGluProGlyLysAlaProGluThrLysGlnGlyThrIleLysValVala	360
Qy	1333	-----GAGCAGAAGTCTACGGAAGACCAACCGGATATAACGGTAGAATT	1374
Db	361	GlnAlaAlaAlaLysLysAspGluLysLysGluAspSerLysLysThrLysThrProVal	380
Qy	1375	GAGAAAGCTGCAAAATCCAAAGTGGCAAGAAGAACTCAGATCTTCCAGCCTGTGATAGAG	1434
Db	381	GluGlu-----HisProLysGlyLysLysGlnGluLysLysGluLysTyrValGluPro	398
Qy	1435	GCGCCTGTGCTCAAAATGTATTGGCCCGCGGTCTGTACAGTGGCGGACGCCCACTCG	1494
Db	399	AlaLysSerSerLysLys-----GluHisSerAlaProSerGluLys	412

Qy	1495	GTFVACTGCGAGTAATGAAGTCTATCCTCAACAC-----GCCGACGCACAAATGAAG	154
Dd	:	:::	
Dd	413	GlnValLysAlaLysThrGluArgAlaLysGluThrSerAlaLaSerThrLysLys	432
Qy	1546	TFTCTAGCTCAGGTAAAGAACAAGAACCCAAAGCCTAAAGAAAAGATGAAGATGAAGCCA	1605
Dd	:	:::	
Dd	433	AlaValProGlyLysLysGluGluLysThrThrllysThrValGluGluLeuAargLys	452
Qy	1606	GAGAAGCCCCAGTCTTCGAAATGGTGCTCAGCAGGTTATTAAATCTCTTCTGTGCAC	1655
Dd	:	:::	
Dd	453	GluLysSergLy-----LysThrSerThrAlaser	462
Qy	1666	AAGAGACCAGCTCCAGAA-----AAAAAGACACACAGCTGAAGAAGCCA-----GTGGTG	1716
Dd	:	:::	
Dd	463	LysAspLysGluProGluLeuLysLysAspGluLysMetProLysAlaAspLysGluVal	482
Qy	1717	GTCCCTCCGCGAGTGAAGCA-----CTCGGAAGAAGACAGCTGTGTGAGAGCAGCACCGCG	1773
Dd	:	:::	
Dd	483	LysProLysProProlnserGlnValLysLysGluLysSerGlnValLys	502
Qy	1774	TCGPGGCGAGCGATCATCAATTACAGTAAAGCCAGAAAAGACTGTGCT	1827
Dd	:	:::	
Dd	503	LysGluAlaLysProGluGln---AspIleAlaLysProGluLysThrValSer	519
<hr/>			
RESULT 14			
RRE1 HUMAN			
ID	RRE1_HUMAN	STANDARD;	PRT; 755 AA.
AC	Q92766;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUL-1998 (Rel. 36, last sequence update)		
DT	10-OCT-2003 (Rel. 42, last annotation update)		
DE	RAS-responsive element binding protein 1 (RREB-1).		
GN	RREB1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	[NCI]		
RW	{1}		
RP	SEQUENCE FROM N.A.		
TX	TISSUE=Thyroid carcinoma;		
RX	MEDLINE=36413283; PubMed=8816445;		
RA	Thiagalingam A., de Bustros A., Borges M., Jasti R., Compton D.,		
RA	Diamond L., Mabry M., Ball D.W., Baylin S.B., Nelkin B.D.;		
RT	"RREB-1, a novel zinc finger protein, is involved in the		
RT	differentiation response to Ras in human medullary thyroid		
RL	carcinomas.";		
RL	Mol. Cell. Biol. 16:5335-5345(1996).		
CC	- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS SPECIFICALLY TO THE		
CC	DISTAL RAS-RESPONSIVE ELEMENT (RRE) IN THE CALCITONIN GENE		
CC	PROMOTER AND AUGMENT THE RAS/RAF-MEDIATED TRANSCRIPTIONAL RESPONSE		
CC	OF THAT PROMOTER. MAY BE INVOLVED IN RAS/RAF-MEDIATED CELL		
CC	DIFFERENTIATION.		
CC	- SUBCELLULAR LOCATION: Nuclear.		
CC	- TISSUE SPECIFICITY: EXPRESSED IN HEART, PLACENTA, LUNG, LIVER,		
CC	SKELETAL MUSCLE, KIDNEY AND PANCREAS. NOT FOUND IN THE BRAIN.		
<hr/>			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
<hr/>			
CC	EMBL; U26914; AAB19094.1; --		
DR	HSP; P08045; IZNF		
DR	TRANSFAC; T01975; --		
DR	Genew; HGNC:10449; RREB1.		
DR	MIN; 602209; --		
DR	GO; GO:0007275; P:development; TAS.		
DR	GO; GO:0007265; P:RAS protein signal transduction; TAS.		
DR	GO; GO:0006366; P:transcription from Pol II promoter; TAS.		

DR InterPro; IPR007087; Znf C2H2.
 DR Pfam; PF00096; Zf-C2H2; 4.
 DR SMART; SM00355; Znf C2H2; 4.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 4.
 KW Transcription regulation; Activator; zinc-finger; Metal-binding;
 XW DNA-binding; Nuclear protein; Repeat.
 FT ZN_FING 315 337 C2H2-TYPE.
 FT ZN_FING 461 483 C2H2-TYPE.
 FT ZN_FING 580 602 C2H2-TYPE.
 FT ZN_FING 608 630 C2H2-TYPE.
 SQ SEQUENCE 755 AA; 79865 MW; 28D863EF62FP8415 CRC64;

Alignment Scores:
 Pred. No.: 0.016 Length: 755
 Score: 173.00 Matches: 138
 Percent Similarity: 33.66% Conservative: 70
 Best Local Similarity: 22.33% Mismatches: 216
 Query Match: 3.69% Indels: 194
 DB: 1 Gaps: 30

US-09-787-016a-1 (1-2610) x RRE1_HUMAN (1-755)

QY	305	AGCCATCAACCCACCAAG	-----AGTTCAGAAACATGGG	346
DB	37	GlnProSerProCysProAlaProGlyProSerLeuProValThrieuGlyProSerGly	56	
QY	347	GTTCGAGGACCACTATCGCAAGCGAGAGGGG	-----	382
DB	57	IleLeuGluSerProMetAlaProAlaProAlaThProGluProProAlaGlnPro	76	
QY	383	---CAGGACCGGAGGCTGACCCACTGGAGCGCCGCCACCCACAGCAGCAGCTGGGCC	439	
DB	77	LeuGlnGlyProValGlnLeu	83	
QY	440	TGTCCCTCGCGCGAGTGGAGGAGCCCAAGCGCATGAGCGTGGAGCAGTTCCTGA	499	
DB	84	-----AlaValProIleTyrSerSerAlaLeuValSerSerPro	96	
QY	500	CCATTGCGGCGCGCGGCA	-----GGAGGAGCATGCTCTCTCCCTGGAGGATT	550
DB	97	ProLeuValGlySerSerAlaLeuLeuSerGlyThrAlaLeuLeuArgProLeuArg	115	
QY	551	CTGGTGAAGCCACGT	-----CCTGCCCGCCACAGACCGCAGACAGCCT	595
DB	116	-----ProLysProProLeuLeuLeuProLysProValThrieuGlyLeuPro	132	
QY	596	CCGAGGCGAGCTGGAAAGCGTCTCAGACCA	-----GAAGCGCGCCCGAGTCTGCTTCCA	652
DB	133	ProLeuAlaSerIleAlaGlnIleIleSerSerValSerSerAlaProThrLeuLeuLys	152	
QY	653	CAGCTGTGAAGAAACGACACGCT	-----	676
DB	153	ThrLysValAlaAspProGlyProAlaSerThrGlySerAsnThrThrAlaSerAspSer	172	
QY	677	CTTCTGAAAGGTGAAGAGGGATGACACGATGACCTCCGATAGTGCAGCGATG	736	
DB	173	LeuGlyGlySerValProLysAlaAlaThrThrAlaThrProAlaAlaThrThrSer	191	
QY	737	GCTGACCTTGAAGAGCTTCAGATCGCTTCGACGAGGAGCGGGAACAGGAGCCACTG	796	
DB	192	-----Pro-LysGluSerSerGluProPro	199	
QY	797	AGAGGCCCTCGAAGGATCCAGAG	-----TGGCTCGGGAAGAAGCCCGGAGGAGG	850
DB	200	-----AlaProAlaSerSerProGluAlaAlaSerProThrGluGlnGlyProAlaAlaArgTh	218	
QY	851	GTCCCGCGAGCTGTGGCTCGAGCGCAGTGCACACTGTGGGGGCGCTCCGCCAGTA	910	
DB	218	-----SerLysLysArgGlyArgLysArgGlyMet	232	
QY	911	AGCAGGAGCCGAGAACGATCAGGGGGTGTGTGCC	-----	945

DB	233	-----ArgAlaAsnSerGlyGly-ValAspLeuAspSerSerGlyGluPheAla	249	
QY	946	-----	-----CAGCGTG	952
DB	249	erIleGlyLysMetLeuAlaThrThrAspThrAsnLysPheSerProPheLeuGlnThrA	269	
QY	953	GGAAGATGACAGAGAGTAAGTTCAGAGGGAAGGCGCTCAGGACATCAAGATGAGG	1012	
DB	269	laGluAspAsnThrGlnAspGluValAlaGly	-----A	280
QY	1013	AGCTGAGAGACTTGGCGCCGACCGAGCTGAATGTGAGGTTACGACCCC	-----AACGCC	1069
DB	280	laProAlaAspHisHisGlyProSerAspGluGlnGlySerProGluAspLysL	300	
QY	1070	TGTATTGATTTGCCGCGAGCTCAACACAGGTTATG	-----ATTGCTGACCGCT	1126
DB	300	euLeuArgAlaLysArgAsnSerTyrThrAsnCysLeuGlnLysIleThrCysProHisC	320	
QY	1127	GTGAAGAATCGTTT	-----CATGCGGATT	1150
DB	320	ysProArgValPheProTyrAlaSerSerLeuGlnArgHis**LeuThrHisThrAsp	340	
QY	1151	GTGGGCGATTCTGAG	-----GCTGAGGAGCGCTTTTGAAGAGGATCGGAG	1201
DB	340	erGlnSerAspAlaGluThrAlaAlaAla**GlyGluValLeuAspLeuThrSerArgA	360	
QY	1202	AC	-----TATATCTGCCAACTGCACCA	1225
DB	360	spArgGluGlnProSerGluGlyAlaThrGluLeuArgGlnValAlaGlyAspAlaProV	380	
QY	1226	TTCTGCAAGTCAGAGTATTCATTCAGAAACGCGAGATCAGCAGGAGAGCTAAA	-----	1281
DB	380	alGluGlnAlaThrAlaGluThrAlaSerProValHisArgGluGluHisGlyArgGlyG	400	
QY	1282	-----TGAGACCTCGAGATCTGATGCGACCGATTGTACAAGTATAGAAACAATAGAGC	1336	
DB	400	luSerHisGluProGluGluGluHisGlyThr	-----GluG	412
QY	1337	AGAACTCTAGCAAGACCAAGGATAAAGGTAGATTTGAGAAGCTGCAAAATCCAGTG	1396	
DB	412	luSerThrGlyAspAlaAspGlyGlyArgGlyArgValGluGln	-----ProGluP	429
QY	1397	GCAAGAGAACTCAAGATCTTCAGCTGTGATAGAGCGCTGCTGCTCAAAATGTA	1456	
DB	429	roGlyProGlyLeuArgHisGlnAlaHisGlyLeuGlnAlaGlyGlyGlyArgArgG	449	
QY	1457	TTGCGCCCGT	1516	
DB	449	lyArgProGlyGlyAla	-----AlaSerGlnGlnGlnLysLeuAlaCysaspThr	465
QY	1517	TCCTCAACACCGCGGACAAATGAATGATTTCTTAAGCTCAGGTAAAGAACAGAGCAA	1576	
DB	466	-----CysGlyLysSerPheLysPheLeuGlyThrLeuSerArgHisArg	481	
QY	1577	AGCTTAAGAAAGATGAAGATGAAGCCAGAGAGCCCGCTTCCGAAATCGCGTGCTC	1636	
DB	481	ysAlaHisGlyArgGlnGluProLysaspGlyLysGlyAsp	-----	494
QY	1637	AGCAGGTATTAAATCTTCTGTGCACAAAGACAGCAGCTCCAGAAAAA	-----	1689
DB	495	-----GlyAlaThrThrAlaGluGluGlyProSerProAlaProGluGlnGluLysP	513	
QY	1690	-----GAGACCACTGAAGAGCGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1744	
DB	513	roProGluThrProAlaGluValValGluSerAlaProGlyAlaGlyGluAlaProAlaG	533	
QY	1745	AGGAGCAGCTGTGAGAGCAGCAGCGCGTGTG	-----	1780
DB	533	luLysLeuAlaGluGluThrGluGlyProSerAspGlyGluSerAlaAlaGlyLysArgS	553	
QY	1781	CGAGCATCAATTCATTCAGTAAGCCAGAAAGACAGCTGCTGCC	1830	
DB	553	erSerGlyLysSerAspAspLysLysPro	-----LysThrAspSerPro	568

QY 580 -----GACGCGAGACACCTCCGAGGCGAGCGTGGAAAGCGCTTCTGAGACGAGA 630
Db 698 AlaGlnAlaGlnValThrAlaProProLeuLysThrAspSerAlaLysThrSerGln 717
QY 631 AGCGCCGCCATCGCTTCCACA-----GCTGTGAAGGAAGACGACA 672
Db 718 SerPheProProThrGlyAspThrIleThrProLeuAspSerLysAlaMetProArgPro 737
QY 673 GCTCTCTCGAAAGCTGAAGAGGAGGATGACACGATGACACCTCCGATAGTGACAGC 732
Db 738 AlaSerAspSerLysIleVal-----SerHisProGlyProThrSerGluSerLys 754
QY 733 GATGCGCTGACCTTGAAGAGCTTCAGAAATCCCTTCGAGGAAGCGGGAACAGAGCGCC 792
Db 755 AspProValGlnLysLysGluLysGluProLysLysAlaGlnThrLysValThrProLysPro 774
QY 793 ACTGAGAGCCCGCTG---AAAGGGATC----- 816
Db 775 AspThrLysProValProLysGlySerProThrProSerGlyThrArgProThrThrGly 794
QY 817 -----CAGAGTCGCTCGGGAAGAGCGCGGAGAGGGT----- 852
Db 795 GlnAlaThrProGlnSerGlnGlnProProLysProProGluLysSerArgArgPheSer 814
QY 853 -----CCGCGCGAGACT 864
Db 815 LeuAsnLeuGlyIleAlaAspAlaProLysSerGlnProThrThrProGlnGluThr 834
QY 865 GTGGGC-----TCCGAGCGCAGTCACTGTG 891
Db 835 ValThrGlyLysLeuPheGlyPheGlyAlaSerIlePheSerGlnAlaSerAsnLeuIle 854
QY 892 GAG-----GGCGTCTCGCCGAGTGAAGCAG 915
Db 855 SerThrAlaGlyGlnGlnAlaProHisProGlnThrGlyProAlaAlaProSerLysGln 874
QY 916 GAGCCGAGACGATCAGGGGTGTGTCCTCCAGGCT----- 951
Db 875 AlaProProSerGlnThrLeuAlaGlnGlyProProLysSerThrGlyGlnHis 894
QY 952 -----GGGAAGATGACAGAGAGT 972
Db 895 ProSerAlaProAlaLysThrThrAlaValLysLysGluThrLysGlyProAlaAlaGlu 914
QY 973 AAGTTGGAGGAAGCGGCTCAG-----GACATCAAGATGAGAG----- 1014
Db 915 AsnLeuGluAlaLysProAlaGlnAlaProThrValLysLysAlaGluLysAspLysLys 934
QY 1015 -----CCTGAGACTGTGGC----- 1029
Db 935 HisProProGlyLysValSerLysProProProThrGluProGluLysAlaValLeuAla 954
QY 1030 -----CGACGAGCTGATGTGAGGTTACGACCCGACGCC 1068
Db 955 GlnLysProAspLysThrLysProLysProAlaCysPro----- 968
QY 1069 CTGTATTGCAATTGCGC-----CAGCCTCACACACAC 1101
Db 969 -----LeuCysArgThrGluLeuAsnValGlySerGluAspProProAsnPheAsn 985
QY 1102 AGGTTTATGATTGC-----TGTGACCGCTGT----- 1128
Db 986 ThrCysThrGluCysLysAsnGlnValCysAsnLeuCysGlyPheAsnProThrProHis 1005
QY 1129 -----GAAGAATGGTTTCATGGCGATTGTGGCATTCTCGAGCTCGAGG 1176
Db 1006 LeuThrGluIleGlnGluThr----- 1012
QY 1177 AGCTTTTGGAAAGGAATGGGAGAGACTATATCTGCCAACTGACCACTCTGCAAGTG 1236
Db 1013 -----LeuCysLeuAsnCys-----GlnThr 1019

QY 1237 CAGGATGAGACTCATTTCAGAAACGGCAGATCAGCAGGAAGCTAAATGAGACCTGGAGAT 1296
Db 1020 GlnArgAlaIleSerGlyGlnLeuGlyAspMetAspLys-----MetProProAla 1036
QY 1297 GCTGATGGCACCGATTCTACAGTATAGGAACAATAGCAGCAGAGTCTAGCGAAGACCAA 1356
Db 1037 SerSerGlyProLysAlaSerProValProAlaProAlaGluProProGln----- 1054
QY 1357 GGGATAAAGGATAGATTGAGAAAGCTGCAAAATCCAAAGTGGCAAGAGAACTCAAGATC 1416
Db 1055 -----LysThrProThrAlaAlaHisAlaLysGlyLysLysGluThrGlu 1070
QY 1417 TTCAG-----CCTGTGATAGAGCGCTCGCTCAAAATGT 1455
Db 1071 ValLysAlaGluThrGluLysGlnIleProGluLysGluThrProSerIleGluLys--- 1089
QY 1456 ATTGGCCCCGGGTCTCTCAGCTGGCGCAGCCCGACTCGGTGTACTCGAGTAATGACTGT 1515
Db 1090 -----ThrProProAlaVal-----AlaThrAspGln 1098
QY 1516 ATCTCTAAACACGCCGCGCAGCGCAATGAAGTTTCTAAGCTCAGGTAAAGAACAGACCA 1575
Db 1099 LysLeuGluLysSerGluValThrLysSerLeuValSerValLeuProGluLysLysPro 1118
QY 1576 AAGCTTAAAGAAAGATGAAGATGAAGCCAGAGAACCCAGTCTTCCGAAATGCGGTGCT 1635
Db 1119 SerGluGluLysAlaLeuProAlaAspLysLysGluLysLysProProAlaAlaGlu 1138
QY 1636 CAGCAGGTATTAAATCTCTCTGTCACACAGACAGCAGCTCCA----- 1680
Db 1139 AlaProProLeuGlu-----GluLysLysProLysProAspAspGlnLysLeu 1154
QY 1681 -----GAAAAAAGAGAGACCCACAGTGAAGAAG 1707
Db 1155 ProProAspAlaLysProSerAlaSerGluGlyGluLysArgAspLeuLeuLysAla 1174
QY 1708 GCAATGTGTGCTCTGCGCGAGTGAAGACTCGGGAAGGAGCA-----GCTTGTGAG 1761
Db 1175 HisValGlnIleProGluGluGly---ProIleGlyLysValAlaSerLeuAlaCysGlu 1193
QY 1762 AGCAGCAGCGCTGCTGGCGAGCGATCACAATTACAATGAGTAAAGCCAGAAAGACT 1821
Db 1194 Gly-----GluGlnGlnProAspThrArgProGluAspLeu 1205
QY 1822 GCTGCTCCCTCGCGCTCACTGTTGTATATAA 1851
Db 1206 ProGlyAlaThrProGlnThrLeuProLys 1215

Search completed: April 28, 2004, 10:55:23
Job time : 70.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2004, 10:46:07 ; Search time 129.5 Seconds
(without alignments)
12718.185 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 crcggtggccgctcgccac.....actttaagatcatatcctg 2610

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p,model -DEV=xlp
-O=/cpn2.1/USPTO spool_p/US09787016/runat 28042004 094329 26919/app query.fasta_1.2759
-DB=SPTRMBL_25 -OFMT=FASTA -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09787016 @CGN 1.1 142 @runat 28042004 094329 26919 -NCFU=6 -ICFU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DSV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTRMBL_25:

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332	7.1	2016	5 Q9VG78	Q9VG78 drosophila

ALIGNMENTS	
RESULT 1	
Q9VG78	PRELIMINARY; PRT; 2016 AA.
ID	Q9VG78
AC	Q9VG78; 2016 AA (Tremblrel. 13, Created)
DF	01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE	CG6525 protein.
GN	SPP OR CG6525.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Berkley;
RX	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton R.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

Q9W352 drosophila
Q9VUB5 drosophila
Q8C3Y1 drosophila
Q9W0T2 drosophila
Q9VUB8 drosophila
Q9W0T1 drosophila
Q7Z7D6 homo sapien
Q818F0 euplores cr
Q7Z2C9 drosophila
Q95YM2 procamburus
Q8C9E9 mus musculus
Q8C9E0 mus musculus
Q7SZE4 brachydanio
Q74508 schizosacch
Q7SZX6 brachydanio
Q81WP2 homo sapien
Q81WP1 homo sapien
Q81WP0 homo sapien
Q81WN9 homo sapien
Q81WN8 homo sapien
Q81WN7 homo sapien
Q86SQ1 homo sapien
Q9BUV3 homo sapien
Q95PU8 chironomus
Q78891 drosophila
Q9W596 drosophila
Q35540 mus musculus
Q28687 oryctolagus
Q91255 petromyzon
Q871Y7 neurospora
Q93291 figu rubrip
Q8VDN7 mus musculus
Q9UQ40 homo sapien
Q9HFW4 ustilago ma
Q9UBA8 homo sapien
Q9UQ35 homo sapien
Q8HZW3 canis famil
Q81PS8 drosophila
Q9VNX6 drosophila
Q99192 mus musculus
P97868 mus musculus
Q9EQJ5 mus musculus
O45407 caenorhabdi

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J.P., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg M.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Raine B.C., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
DR EMBL: A5003695; AAF54807.1; --
DR FlyBase: FBgn0038041; Spp.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR006576; BRK.
DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00628; PHD; 1.
DR SMART: SM00592; BRK; 1.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS01359; ZF_PHD_1; 1.
DR PROSITE: PS50016; ZF_PHD_2; 1.
DR SQ SEQUENCE 2016 AA; 221998 MW; PB61A332BA710F5E CRC64;

Alignment Scores:
Pred. No.: 1,47e-14 Length: 2016
Score: 332.00 Matches: 159
Percent Similarity: 32.52% Conservative: 68
Best Local Similarity: 22.78% Mismatches: 221
Query Match: 7.08% Indels: 250
DB: 5 Gaps: 27

US-09-787-016A-1 (1-2610) x Q9VG78 (1-2016)
QY 256 GGAGCAGGATGACGACCAAGCGGACCGAGCAATGAGGAGCACTAAGGCCATCAA 315
Db 676 GlyValGlyAsnAlaAlaAlaAspProAsn-----ileMet 689
QY 316 CCCACAGCAAGAGTTCAGGAAACATGGGTTTCGAAAGGACCCTATCGCCAGCGA 375
Db 690 AspThrAlaAsnGluAspGluIleThrAlaAspPheLeuGlnHisValValGlyLeulle 709
QY 376 GAGGCGCGAGCGGAGCGAGGTGACCCACTGGAGCCGCCACCCACAGCAGGAGCTG 435
Db 710 GluGluAspIysGlnPheGluAla----- 717
QY 436 GGCCTGTCCCTGGCGGCGAGTGGGAGGACCCCAAGCGCACTGAGCGCGTGGAGGATTC 495
Db 718 -----GluValValIysGlnVal 723
QY 496 CTGACCATTCGCGGCGCGCGCAGGAGGAGCATGCTCTCCCTGGAGGATTCGTGT 555
Db 724 LeuAla----SerThrGluProGlyThrLeuAspAlaIleValSerMetProThrSerIle 742

QY 556 GAGCCCAAGTCTGCTGCGCCGACACGACCGAGAGCAGCTCCGAGGGGAGCTG---GAA 612
Db 743 GluProValAspValProGlnAlaHisThrAsnLeuLeuProAsnAlaSerLeuThrGlu 762
QY 613 ACCGCTTCTGAGACAGAGAGGGGCCCGCTGCTGCTTCCACA-----GCT 657
Db 763 ProAlaGlnSerMetThrSerLeuProIleAlaCysSerThrProSerArgSerValAla 782
QY 658 GTGAGGAAACGACCAAGCTCTTCTGAAAAGGTGAAGAGGGGATGACCAACGATGACACC 717
Db 793 AlaSerThrProProThrSerAlaIysValValArgGly----- 795
QY 718 TCCGATAGTACAGCGATGGCTTGAAGAGAGCTTCAAGATCGCTTCGAGGAGAG 777
Db 796 -----TyrglyArgVal 799
QY 778 CGGGAACAGGAGCCCACTGAGAGGCCCTCTGAAAGGATCCAGAGTGGCTCGGGAAGAG 837
Db 800 IleTyLeuProProIleGluAlaPro-----ThrThrArgAlaIys 813
QY 838 CGCCGGGAGGAG-----GGTCCCGCCGAGACTGTGGGTTCGAGGCC--- 879
Db 814 ArgArgAlaGlnPheProSerAlaProGlyMetAlaAlaThrSerSerSerAspAlaGly 833
QY 879 ----- 879
Db 834 AsnLeuSerPheGlyGluSerSerLeuAspAlaSerIleAsnGlnProLeuAsnThrSer 853
QY 880 -----AGTGACACTGTGGAGGGCTCTGCCCCTAAGCAGGAGCCCGAGAAAGAT 930
Db 854 SerLeuSerAsnAspSerGlnProGlySerGlyProLysArgProAsnPro---ArgGlu 872
QY 931 CAGGGGGTGTGTCGCCAGGCTGGGAAAGATGACAGAGAGTAACTGAGGGGAAGGGCG 990
Db 873 ProSerMetAlaArgArgSerThrAlaProArgArgSerLysLysLeuAspAlaSerGln 892
QY 991 GCTCAGGAC-----ATCAAGATGAGGAGCTTGGGAGCTTGGCCGACCGAAGCCT 1041
Db 893 AsnAsnAspProAspAlaSerGluSerGlnGluAspAsp----- 906
QY 1042 GAATGTGAGGTACGACCCCAAGCCCTGTATTCATTTGCCCTGCGGAGCTCAGACAC 1101
Db 907 -----AspProAsnLysLeuTrpCysIleCysArgGlnProHisAsnAsn 921
QY 1102 AGGTTTATGATTGCTGACCGCTGTGAAGATTCGTTTCATGCGCATTTGTGGGCATT 1161
Db 922 ArgPheMetIleCysCysAspLeuCysGluAspTrpPheHisGlyThrCysValGlyVal 941
QY 1162 TCTGAGGCTCGAGGAGGCTTTTGAAGAGAAATGGGGAAGACTATATTCGCCCAACTGC 1221
Db 942 ThrLysAlaMetGlyThrAspMetGluAsnLysGlyIleAspTrpLysCysProLysCys 961
QY 1222 ACCATTCTCAAGTCAGGATGAGACTCATTCAAGAACGGCAGATCAGCAGGAAGCTAA 1281
Db 962 -----ValLysArgGlnGluArg 968
QY 1282 TGGAGACCTGGAGATGCTGATGGCCCGATTTGTACAGTATAGGAACAATAGAGCAGAG 1341
Db 969 SerGlnProArgIleThrAspMetLeuValThrArgProThrThrGlnProGluArg 988
QY 1342 TCTACGAGAGACCAAGGATA-----AAGGTAGAAATGAGAAAGCTCCAAATCCAAGT 1395
Db 989 ProSerGluThrLysValLeuThrThrAlaGluIleValGlnValAlaAlaProSer 1008
QY 1396 GGC---AAGAAGAACTCAAGATCTCCAGGCTGTGATAGAGGCGCCT----- 1440
Db 1009 AlaProArgArgThrLeuProValValLeuThrValAlaSerSerProMetArgIlePro 1028
QY 1441 -----GGTGCCTCAAAA----- 1452
Db 1029 MetAlaLysProAlaLysPheProThrGlyAlaIleSerHisGlnGlnGlnGln 1048

Db 87 AspCysIleGlyIleThrGluLysGluAlaLysHisIleLysGln-----Tyr 102
 QY 1207 ATCTGCCCAAACTGC-----ACCATCTCGCAA----- 1233
 Db 103 TyrCysArgCysLysLysGluAsnProGluLeuGlnThrIlePheArgLeuValAla 122
 QY 1233 ----- 1233
 Db 123 ThrGluArgAlaAlaAlaSerAsnAlaAlaSerThrSerLeuAsnAlaProGlyValGly 142
 QY 1234 -----GTCGAGGATGAGACTCATTTCAGAAACCGCA 1263
 Db 143 ProSerGlyAlaAlaProAlaAlaAlaProValAlaSerAlaThrThrSerGlnGlnAla 162
 QY 1264 -----GATCAGCAGAGCACTAAATGG 1284
 Db 163 ProProProThrThrAlaAlaAlaLysArgLysAsnSerSerAlaArgGluProLysMet 182
 QY 1285 -----AGACCTGGAGATCGCTGATGC----- 1305
 Db 183 GlyLysArgCysGlyThrCysGluGlyCysArgArgProAsnCysAsnGlnCysAspAla 202
 QY 1306 -----ACCATTTGTTCAAGTATA 1323
 Db 203 CysArgValArgValGlyHisLysProArgCysIlePheArgThrCysValValGlnAla 222
 QY 1324 GGAACATAGACGAGAGTCTAGCGAGACCAA---GGATTAAGCGTAGATTTCAGAAA 1380
 Db 223 AlaThrValLeuLysGluSerGlnAlaThrGlnAlaGlyProSerArgLysArgGluLys 242
 QY 1381 GGTGCAATCCAAAGTGGCAAGAAAGAACTC-----AAGATCTCCAGCCTGTGATA--- 1431
 Db 243 AlaAlaProLysSerArgAsnValGlnValGlyProArgAlaAlaSerProGluIlePhe 262
 QY 1432 -----GAGCGCTGCTGCTCAATATGTTGGCCCGGTGCTGTCACGTGGCG 1482
 Db 263 LeuAsnProGluGlnGlyIleArgGlnCysTyrGlyProAsnCysCysSerHisAla 282
 QY 1483 CAGCCCGCATCGGTGACTGCAAGTACTGCT----- 1515
 Db 283 ArgProGlnSerLysTyrCysSerAspLysCysGlyPheAsnLeuAlaThrLysArgIle 302
 QY 1515 ----- 1515
 Db 303 PheGlnValLeuProGlnArgLeuGlnGluTrpAsnLeuThrProSerArgAlaAlaGlu 322
 QY 1516 -----ATCTCAACACGCGCGCAGCACAATGAAGTTTCTAAGC 1554
 Db 323 GluThrArgLysHisLeuAspAsnIleArgHisLysGlnSerLeuValArgPheAlaLeu 342
 QY 1555 TCAGGTAAGACAGAACGCAAGCTTAAGAAAGATGAAGATGAGACGACGAGAGCCG 1614
 Db 343 AlaGluLeuGluLysArgSerGluGluLeuAsnMetValValGluArgAlaLysArgSer 362
 QY 1615 AGTCTCCGAATCGGTGCTCAGCGAGGTATTAATCTCT----- 1656
 Db 363 SerIleAspThrLeuGlySerGlnAlaAspThrAlaAspMetGluAspGluGlnSerMetTyr 382
 QY 1657 -----TCTGTGCACAGACACGACCTCCAGAA----- 1683
 Db 383 CysIleThrCysGlyHisGluIleHisSerArgThrAlaIleLysHisMetGluLysCys 402
 QY 1684 -----AAAAAGACACACAGTACGAGAGGCGAGTGTGCTCGCTCGCGGAGTACGACA 1737
 Db 403 PheAsnLysTyrGluSerGlnAlaSerPheGlySerIlePheLysThrArgMetGlu--- 421
 QY 1738 CTCGGTAAGACGACGCTGTGACAGCAGCAGCGCGCTCG 1776
 Db 422 ---GlyAsnAsnMetPheCysAspPheTyrAsnProAla 433

RESULT 3

Q9VUB5

PRELIMINARY; PRT; 3146 AA.

RP

SEQUENCE FROM N.A.

Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,

AC

Q9VUB5;

01-MAY-2000 (TrEMBLrel. 13, Created)

01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

CG9007 protein.

CN

Drosophila melanogaster (Fruit fly).

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

RN

SEQUENCE FROM N.A.

PC

STRAIN=Berkeley;

MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M.R., Miklos G.L.G.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Baldwin D.,

RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,

RA Fofor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,

RA Glöde K., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houch J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Neelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirska R.A., Tector C., Turner R., Venter B., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

"The genome sequence of Drosophila melanogaster."

Science 287:2185-2195(2000).

RN

SEQUENCE FROM N.A.

RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,

RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,

RA Banon J., An H., Baldwin D., Banon J., Beeson K.Y., Busam D.A.,

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,

RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,

RA Ferrieria S., Frise E., Galle R.F., Garg N.S., George R.A.,

RA Gonzalez M., Houch J., Hoskins R.A., Hostin D., Howland T.J.,

RA Ibegwan C., Jallali M., Ke Z., Kennison J.A., Ketchum K.A.,

RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,

RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,

RA Phoenanavong S., Pittman G.S., Puri V., Richards R., Tector C.,

RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,

RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,

RT "Sequencing of Drosophila melanogaster genome."

Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN

[3]


```

QY 1327 ACAATA-----GAGCAGAAAGTCTAGC 1347
DB 939 ThrThrLeuSerGlyGlyLeuGlySerGlyLeuProMetSerGluGluLeuGlnHisArg 958
QY 1348 GAGACACAGGATAAAGGTAGATT----- 1374
DB 959 LeuAlaSerGlyLeuAsnGlyGlyPheAlaThrGlyThrGlyMetSerLysLysSerLys 978
QY 1375 -----GAGAAAGCTCAATCAAGTGGCAGAGAAACTCAAGATCTTCACGCT 1425
DB 979 LysThrLysGluAsnSerGlySerThrSerThrLeuLysLysThrLys-----LysSer 996
QY 1426 GTGATAGAGCGCGTGGTGGCTCAAAATGATTGGCCCGGGTGTGTCACTGGCGGAG 1485
DB 997 AlaValGlyMetGlyGlyGlyLysAsnAlaSerGlySerGly-----Thr 1011
QY 1486 CCGAGCTCGGTGACTGATGATGATCTCTCAACACAGCCGCGCAATGAAG 1545
DB 1012 ProThrGly----- 1014
QY 1546 TTTCTAAGCTCAGTTAAGACAGAACGCAAGCCTTAAGAAAGATGAAGATGAAGCCA 1605
DB 1015 -----SerSerGlyLysThrSerLysLysSerSerLysArgLysSerLysSerGlyGly 1032
QY 1606 GAGAACCCAGCTTCCGAAATCGGTGCTCAGGCAGGTATTAAATCTCTCTGTGCGAC 1665
DB 1033 Asp-----GlySerSerGlyGly-----GlySerSerPro 1042
QY 1666 AAGAGACAGCTCCAGAAAAAAGAGACACAGTGAAGAGCAGTGTGTGCTCGCTGG 1725
DB 1043 AlaLeuThrAlaAlaGlyLysHisAlaAlaAsnLeuArgGln----- 1056
QY 1726 CGGAGTGAAGCACTCGGGAAGGAGCAGCTGTGAGAGCAGCAGCGCTGGCGGAGC 1785
DB 1057 -----TrpIleGlu 1059
QY 1786 GATCACAAATAC-----AATGCAGTAAGCCAGAAAGACTGCT----- 1824
DB 1060 AsnTyrGluTyrAlaValThrAsnHisTyrSerProGluLeuArgAlaGluHisAla 1079
QY 1825 ----GCTCCCTCGCGTCACTGTTG 1845
DB 1080 IleGlnLysGlnProSerLeuLeu 1087

RESULT 4
ID Q8T3Y1 PRELIMINARY; PRT; 366 AA.
AC Q8T3Y1; Q9W353;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE AT26187p (CG17440 protein).
GN CG17440
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacle J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196066; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

```

```

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbavani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Fostler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirska R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY089443; AAL90181.1;
DR

```

DR EMBL; AE003446; AAP46482.2; --
DR FlyBase; FBgn0030120; CG17440.
DR GO; GO:0003577; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001965; znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
SQ SSQUENCE 366 AA; 42048 MW; CB83AAPFAACS14CF3 CRC64;

Alignment Scores:
Pred. No.: 1,07e-05 Length: 366
Score: 202.50 Matches: 63
Percent Similarity: 37.55% Conservative: 29
Best Local Similarity: 25.71% Mismatches: 82
Query Match: 4.32% Indels: 71
DB: Gaps: 8

US-09-787-016A-1 (1-2610) x Q8T3Y1 (1-366)

QY 910 AAGCAGGAGCCCGAGACGATCATGGGGGTGTGTGCCAGGCTGGGAAGAATGCACAGAG 969
Db 7 LysThrGluGluIleArgArgGluIleAlaargLupheaspLeuProGluarGLys 26
QY 970 AGTAAGTTGGAGGGAAGAGGGGCTCAGGACATCAAGATGAGGAGCCCTGGAGACTTGCGC 1029
Db 27 SerLysilealaThrlieLeuLysglN-----LysAspargLu----- 39
QY 1030 CGRCCGAGGCTGAATTGAGGGTTACGACCACCCCTGTATTTCATTTGCCGCCAG 1089
Db 40 -----TyrcyllieCysArSer 45
QY 1090 CCTCACAAACACAGGTTTATGTTGTCACCGCTGTGAAGAANGTTTCATGGCGAT 1149
Db 46 SeraspCysSerArgPheMetIleGlyCysaspGlyCysGluGluTptYrHisGliVasp 65
QY 1150 TGTGTGGCATTTCTGAGGCTCGAGGGAGGCTTTTGGAAAGGAATGGGAGACTATATC 1209
Db 66 CyeileGluileThrGluLysaspAlaGluHifle-----LysAsnTYr 81
QY 1210 TGCCCCAACATGC-----ACCATTCGCAAGTCCAGGATGAG 1245
Db 82 CysargArgCysLysLysGluasnProGluLeuGlnThrillePheArgLeu----- 98
QY 1246 ACTCATTCAGAACCGCAGATCAGCAGGAGCTAAATGGAGACCTGGAGATGCTGATGCG 1305
Db 99 ValalathrGluArgAlaAlaalaserAsnalAlaserThrserLeuAsnalAlaproGly 118
QY 1306 ACCGATTGTACAAGATATAGGAACAATA----- 1332
Db 119 ValGlyProserGlyAlaAlaProAlaalAlaProValAlaProAlaThrThrSerGln 138
QY 1333 -----GAGCAG 1338
Db 139 GlNALaProProthrThrAlaAlaalalysArgLysasnSerSerAlaGlnGluPro 158
QY 1339 AAGTCTAGCAAGACCAAGGATAAAGGGTAGAATTCAGAAAGCTCGAAATCCAAGTGGC 1398
Db 159 LysGluSerGlnProThrGlnAlagLyThrLysArgaspLysalaAlaProLysThrSer 178
QY 1399 AAGAGAAACTC-----AGATCTTCACGCTGTGATA-----GAGCGCGCT 1440
Db 179 AsnValGlnValserProArgAlaValiserProGluilePheLeuAsnProGluLeuGln 198
QY 1441 GGTCCTCAAATATATTGGCCCGGCTGTGTCCGTCGGCAGCCGACTCGGTGTAC 1500
Db 199 GlytleGlnGlnCynHisGlyProAsnCysCysSerHisalaaArgProGlnSerLysTYr 218
QY 1501 TGCAGTAATGACTGT 1515
Db 219 CysSerAspGluCys 223

RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Carlson J.W., Celisner S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (XAR-2000) to the EMBL/GenBank/DBSJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF003467; AAF47361.2; -;
DR FlyBase; FBgn000541; E(hx).
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003489; F:electron transporter activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR006337; AT hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC heme_BS.
DR InterPro; IPR004022; DDT dom.
DR InterPro; IPR006289; EGF-like.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT hook; 1.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF06528; PHD; 3.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 3.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00114; BROMODOMAIN_2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2649 AA; 298507 MW; CODF7E015EA5403 CRC64;

Alignment Scores:
Pred. No.: 3.13e-05 Length: 2649
Score: 198.50 Matches: 87
Percent Similarity: 36.75% Conservative: 78
Best Local Similarity: 19.38% Mismatches: 147
Query Match: 4.23% Indels: 139
DB: 5 Gaps: 16

US-09-787-016a-1 (1-2610) x Q9W0T2 (1-2649)
QY 362 CTATGCCAGCGAGGGCGGAGGCGGAGCGTGCACCTGAGCGCCACCC- 420
Db 2180 IleThrThrAsnGluSerPheAlaGlyThrSerSerLeuLeuGluGlySerGluHisAsp 2199
QY 421 ---CCACAGCAGCAGCTGGCGCTCTCCCTCGCGCGAGTGGGAGGAGCCCAAGCGCACT 477
Db 2200 GluProThrAsnLeuAlaGlyLeuAspIleSerGluThrAspLeuGluAsnLysGlnAsn 2219
QY 478 GAGCGC-----GTGAGCAGCTTCCTGACCATTCGAGGCGCGCGCGC 516
Db 2220 GluSerPheValThrArgGlyTyrLeuGlnLysSerIleSerAsnAlaLeuLysGln 2239
QY 517 GGACGAGGAGCAGTCGCTGCTCCCTGGAGCT-----TCTGGTGGAG 558
Db 2240 Gly-----AsnLeuSerProGluLeuGluGluLysLeuValCysMetGlnLysGlnGln 2257

QY 559 CCCACGTCCTCCCGCCGACAGACGCGGAGCAGCCTCCGAGGCGCAGCGTGGAAGCGCT 618
Db 2258 GluAsnAlaAsnSerThrAsnGluThrGluThrCysSerArgGlySerValAsn----- 2275
QY 619 TCTGAGACGAGAACGCGCCCGCTGCTCCACAGCTGTGAGGAGAACGACGACCTCT 678
Db 2276 -----GluGluAlaLeuThrProSerArgGlnThrAspAspThr 2288
QY 679 TCTGAAAGAGTGAAGGAGGAGTACGACGATGACACCTCCGATAGTACGACGAGTGGC 738
Db 2289 GluTrpLysIleArg-----ThrSerLeuArgArgProAsnAla 2301
QY 739 CTGACCTTGAAGAGCTTTCAGATCGCTTCGACGAGAGCGGGAACAGGACCCACTGAG 798
Db 2302 MetThrThrSerSerGlnPheAsnArgIleLeuLysLysAsnArgSerLysAsnAspGlu 2321
QY 799 AGGCGCTGAAAGGAGTCCAGAGTCGCTGCGGAGGAGCGCGGAGGAGCGTCCCGCC 858
Db 2322 ValAlaGluLeuGlyGluGlnLysGlnSerGlnLeuGluArgHisLysGluLeuLys 2341
QY 859 GAGACTGTG-----GGCTCGGAGCCGAGTGC 885
Db 2342 LysAsnIleLeuArgLysArgSerLeuLeuGluArgAsnLeuGlnSerGluIleHisGlu 2361
QY 886 ACTGTGGAGCGCTCCCGCTGACGAGTACGAGGCGCGGAGACGATCAGGCGGTGTGTCC 945
Db 2362 AspValLysThrLysValGlnArgHisValArgPro-----LeuSer 2375
QY 946 CAGGCTGGGAAAGATGACAGA---GAGAGTAAGTTGGAGGAGAAAGCGGCTCAGACATC 1002
Db 2376 AsnAlaSerProAspGluGlnSerGluAsnGluArgSerGlyGluProAsnLeuAspPhe 2395
QY 1003 AAAGATGAGGAGCTCGGAGC-----TTGGGCGGACCGGAGCCT- 1041
Db 2396 LysArgThrGluValGlnAsnProArgHisGlyAlaGlyArgProLysLysLeuThrArg 2415
QY 1041 ----- 1041
Db 2416 LysLysGluLysLeuTyrCysIleCysArgThrProTyrAspAspThrLysPheTyrVal 2435
QY 1041 ----- 1041
Db 2436 GlyCysAspLeuCysSerAsnTrpPheHisGlyAspCysValSerIleThrGluGluAla 2455
QY 1042 -----GAATGTGAG---GGTTACGACCCCAACGCC 1068
Db 2456 SerLysLysLeuSerGluPheIleCysIleAspCysLysArgAlaArgGluThrGlnGln 2475
QY 1069 CTGTATTCATTTGCCGCGACCTCACAC---AACAGTTTATGATTTGCTGTCACCGC 1125
Db 2476 LeuTyrCysSerCysArgGlnProTyrAspGluSerGlnPheTyrIleCysCysAspLys 2495
QY 1126 TCTGAGAAATGCTTTCATGCGGATTTGCTGGGCTTCTGAGGCTCGAGGAGGAGCTTTTG 1185
Db 2496 CysGlnAspTrpPheHisGlyArgCysValGlyIleLeuGlnSerGluAlaGluPheIle 2515
QY 1186 GAAAGGAATGGGAGACTATATCTGCCAACTGCACCATTTCTCAAGTCGAGGATGAG 1245
Db 2516 -----AspGluTyrValCysProGluCysGln----- 2524
QY 1246 ACTGATTGAGAAACGCGAGATCAGCAGGAGCTAATGAGACCTGGAGACTCTGATGCG 1305
Db 2525 -----ArgLysAsnAspAlaAsnAla 2531
QY 1306 ACCGATTGTACAAGTATAGGAACAATAGACAGAGTCTAGCGAAGACCGAGGATAAG 1365
Db 2532 AlaAsnMetLysLysLeuThrSerAsnAspValGluGluLeuLysAsnLeuIleLysGln 2551
QY 1366 GTAGAAATGAGAAAGCTGCA-----AATCCAGTGGCAGAGAGAACTCAAG 1413
Db 2552 MetGlnLeuHisLysSerAlaTrpProPheMetGluProValAspProLysGluAlaPro 2571
QY 1414 ATCTTCCAGCTGTGATAGAGCGGCT 1440

478 GAGCGC-----GTGAGCAGGTTCTCTGACCATTCGCGCGCGCGC 516
 2240 GluSerPheValThrArgGlyTyrIleGlnLysAlaIleSerAsnAlaLeuLysGln 2259
 517 GCACGAGGAGCATCCTGCTCTCCCTGAGGAT-----TCTGGTGG 558
 2260 Gly-----AsnLeuSerProGluLeuGluLysLeuValCysMetGlnLysGlnGln 2277
 559 CCCACGTCCTGCCCGCCACAGAGCCGAGAGCCTCCGAGGCGAGCTGGAAGCGCT 618
 2278 GluAsnAlaAsnSerThrAsnGlnTrpGluThrCysSerArgGlySerValAsn----- 2295
 619 TCTGAGACCAAGAGCGCGCCCGAGCTCTCTCCACAGCTGTGAAGGAAGACGACGCTCT 678
 2296 -----GluGluAlaLeuThrProSerArgGlnThrAspAspThr 2308
 679 TCTGAAAGGTGAAGAGGAGGAGGATGACCATGACCTCCGATAGTACAGAGCGATGCG 738
 2309 GluTrpLysIleArg-----ThrSerLeuArgArgProAsnAla 2321
 739 CTGACCTTGAAGAGCTTTCAGAACTCCCTTCGAGGAAGCGGGAACAGAGGCGCACTGAG 798
 2322 MetThrThrSerSerGlnPheAsnArgIleLeuLysLysAsnArgSerLysAsnAspGlu 2341
 799 AGGCCCTGAAGAGGATCCAGAGTCGCTCGAGAGAGAGCGCGGAGGAGGAGGTCGCCGC 858
 2342 ValAlaGluLeuGlyGluGlnLysGlnSerGlnLeuGluArgHisGlySerGlnLeuLys 2361
 859 GAGACTGTG-----GGCTCCGAGGCGCACTGAC 885
 2362 LysAsnIleLeuArgLysArgSerLeuLeuGluArgAsnLeuGlnSerGluIleHisGlu 2381
 886 ACTGTGAGGCGGCTCTGCGCCAGTAAAGAGGAGCGCGAGAGATCAGGGGGTGTGTGCC 945
 2382 AspValLysThrLysValGlnArgHisValArgPro-----LeuSer 2395
 946 GAGGCTGGGAAGATGACAGA---CAGAGTAACTTGGAGGGAAGCGCGCTCAGGACATC 1002
 2396 AsnAlaSerProAspGluGlnSerGlnAsnGluArgSerGlyGluProAsnLeuAspPhe 2415
 1003 AAAGATGAGGAGCTGTGAGAC-----TTGGCGCGACCGAAGCCT----- 1041
 2416 LysArgThrGluValGlnAsnProArgHisGlyAlaGlyArgProLysLysLeuThrArg 2435
 1041 ----- 1041
 2436 LysLysGluLysLeuTyxCysIleCysArgThrProTyAspThrLysPheTyrVal 2455
 1041 ----- 1041
 2456 GlyCysAspLeuLysSerAsnTrpPheHisGlyAspCysValSerIleThrGluGluAla 2475
 1042 -----GATGTGAG-----GGTACGACCCCAACGCC 1068
 2476 SerLysLysLeuSerGluPheIleCysIleAspCysLysArgAlaArgGluThrGlnGln 2495
 1069 CTGTATTCATTTGCGCGAGCTCACAAC---AACAGTTTATGATTTGCTGTGACCGC 1125
 2496 LeuTyxCysSerCysArgGlnProTyAspGluSerGlnPheTyrIleCysCysAspLys 2515
 1126 TGTGAAGATGTTTCATGCGGATTTGTGGCATTTCTGGAGCTTCGAGGAGGAGGCTTTTG 1185
 2516 CysGlnAspTrpPheHisGlyArgCysValGlyIleLeuGlnSerGluAlaGluPheIle 2535
 1186 GAAGGATGGGAGAGTATATCTGCCCAACTCACCATTCTGCAAGTGCAGGATGAG 1245
 2536 -----AspGluTyrValCysProGluCysGln----- 2544
 1246 ACTCATTGAGAAACGGCAGATCAGCAGGAAGCTAAATGAGACCTGGAGATGCTGATGCG 1305
 2545 -----ArgLysAsnAspAlaAsnAla 2551
 1306 ACCGATTGTACAGTATATAGGAACAATAGACGAGAGTCTAGCGAGACCAAGGATTAAG 1365

QY 478 GAGCGC-----GTGAGCAGGTTCTCTGACCATTCGCGCGCGCGC 516
 Db 2240 GluSerPheValThrArgGlyTyrIleGlnLysAlaIleSerAsnAlaLeuLysGln 2259
 QY 517 GCACGAGGAGCATCCTGCTCTCCCTGAGGAT-----TCTGGTGG 558
 Db 2260 Gly-----AsnLeuSerProGluLeuGluLysLeuValCysMetGlnLysGlnGln 2277
 QY 559 CCCACGTCCTGCCCGCCACAGAGCCGAGAGCCTCCGAGGCGAGCTGGAAGCGCT 618
 Db 2278 GluAsnAlaAsnSerThrAsnGlnTrpGluThrCysSerArgGlySerValAsn----- 2295
 QY 619 TCTGAGACCAAGAGCGCGCCCGAGCTCTCTCCACAGCTGTGAAGGAAGACGACGCTCT 678
 Db 2296 -----GluGluAlaLeuThrProSerArgGlnThrAspAspThr 2308
 QY 679 TCTGAAAGGTGAAGAGGAGGAGGATGACCATGACCTCCGATAGTACAGAGCGATGCG 738
 Db 2309 GluTrpLysIleArg-----ThrSerLeuArgArgProAsnAla 2321
 QY 739 CTGACCTTGAAGAGCTTTCAGAACTCCCTTCGAGGAAGCGGGAACAGAGGCGCACTGAG 798
 Db 2322 MetThrThrSerSerGlnPheAsnArgIleLeuLysLysAsnArgSerLysAsnAspGlu 2341
 QY 799 AGGCCCTGAAGAGGATCCAGAGTCGCTCGAGAGAGAGCGCGGAGGAGGAGGTCGCCGC 858
 Db 2342 ValAlaGluLeuGlyGluGlnLysGlnSerGlnLeuGluArgHisGlySerGlnLeuLys 2361
 QY 859 GAGACTGTG-----GGCTCCGAGGCGCACTGAC 885
 Db 2362 LysAsnIleLeuArgLysArgSerLeuLeuGluArgAsnLeuGlnSerGluIleHisGlu 2381
 QY 886 ACTGTGAGGCGGCTCTGCGCCAGTAAAGAGGAGCGCGAGAGATCAGGGGGTGTGTGCC 945
 Db 2382 AspValLysThrLysValGlnArgHisValArgPro-----LeuSer 2395
 QY 946 GAGGCTGGGAAGATGACAGA---CAGAGTAACTTGGAGGGAAGCGCGCTCAGGACATC 1002
 Db 2396 AsnAlaSerProAspGluGlnSerGlnAsnGluArgSerGlyGluProAsnLeuAspPhe 2415
 QY 1003 AAAGATGAGGAGCTGTGAGAC-----TTGGCGCGACCGAAGCCT----- 1041
 Db 2416 LysArgThrGluValGlnAsnProArgHisGlyAlaGlyArgProLysLysLeuThrArg 2435
 QY 1041 ----- 1041
 Db 2436 LysLysGluLysLeuTyxCysIleCysArgThrProTyAspThrLysPheTyrVal 2455
 QY 1041 ----- 1041
 Db 2456 GlyCysAspLeuLysSerAsnTrpPheHisGlyAspCysValSerIleThrGluGluAla 2475
 QY 1042 -----GATGTGAG-----GGTACGACCCCAACGCC 1068
 Db 2476 SerLysLysLeuSerGluPheIleCysIleAspCysLysArgAlaArgGluThrGlnGln 2495
 QY 1069 CTGTATTCATTTGCGCGAGCTCACAAC---AACAGTTTATGATTTGCTGTGACCGC 1125
 Db 2496 LeuTyxCysSerCysArgGlnProTyAspGluSerGlnPheTyrIleCysCysAspLys 2515
 QY 1126 TGTGAAGATGTTTCATGCGGATTTGTGGCATTTCTGGAGCTTCGAGGAGGAGGCTTTTG 1185
 Db 2516 CysGlnAspTrpPheHisGlyArgCysValGlyIleLeuGlnSerGluAlaGluPheIle 2535
 QY 1186 GAAGGATGGGAGAGTATATCTGCCCAACTCACCATTCTGCAAGTGCAGGATGAG 1245
 Db 2536 -----AspGluTyrValCysProGluCysGln----- 2544
 QY 1246 ACTCATTGAGAAACGGCAGATCAGCAGGAAGCTAAATGAGACCTGGAGATGCTGATGCG 1305
 Db 2545 -----ArgLysAsnAspAlaAsnAla 2551
 QY 1306 ACCGATTGTACAGTATATAGGAACAATAGACGAGAGTCTAGCGAGACCAAGGATTAAG 1365

Db 2572 AspTyrTrpLysValIleLysGluPro 2580
 Q95VB8 PRELIMINARY; PRT; 2669 AA.
 AC Q95VB8;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Nucleosome remodeling factor large subunit NURF301.
 GN E(BX) OR NURF301 OR CG7022 OR CG10894 OR CG17135 OR CG32346 OR CG32478.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21468388; PubMed=11583616;
 RA Xiao H., Sandatzopoulos R., Wang H., Hamiche A., Ranallo R., Lee K.,
 RA Fu D., Wu C.;
 RT "Dual functions of largest nurf subunit nurf301 in nucleosome sliding
 RT and transcription factor interactions";
 RL Mol. Cell 8:531-543 (2001).
 DR EMBL; AF417921; AAL16644.1; --
 DR FlyBase; FBgn000541; E(bx).
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0006777; F:DNA binding; IEA.
 DR GO; GO:0005489; F:electron transport activity; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000637; At hook.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR000345; CycC_heme_BS.
 DR InterPro; IPR004022; DDT_dom.
 DR InterPro; IPR006209; EGF-like.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF02178; AT hook; 1.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF00628; PHD; 3.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00249; PHD; 3.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 1.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS01359; ZF_PHD_2; 2.
 DR PROSITE; PS50016; ZF_PHD_2; 2.
 SQ SEQUENCE 2669 AA; 300687 MW; 654925AFPF489D6F CRC64;

Alignment Scores:
 Pred. No.: 3.13e-05 Length: 2669
 Score: 196.50 Matches: 87
 Percent Similarity: 36.75% Conservative: 78
 Best Local Similarity: 19.38% Mismatches: 147
 Query Match: 4.23% Indels: 138
 DB: Gaps: 16

US-09-787-016A-1 (1-2610) x Q95VB8 (1-2669)
 QY 362 CTATCCCAAGCAGAGGCGGAGCGGAGCGGAGCTGACCCACTGGAGCGCGCACCC- 420
 Db 2200 IleThrThrAsnGluSerPheAlaGlyThrSerSerLeuLeuGluGlySerGluHisAsp 2219
 QY 421 ---CCACAGCAGCAGCTGGCGCTGCTCCCTCGCGCAGTGGGAGGCGGCCAAGCGCACT 477
 Db 2220 GluProThrAsnLeuAlaGlyLeuAspIleSerGluThrAspLeuGluAsnLysGlnAsn 2239


```

Db      2422 leuAspLysIleAspLysGluGluGlnAlaAlaLysLysArgLysArgGluGlu 2442
QY      854 CCGCCGAGACTGTGGCTCCGAGGCGAGTGCACACTGTGGAGGCGCTCCGAGTAAAGC 913
Db      2442 ervAlGluGlnLysArgSerLysGlnAsnAlaThrLysLysSerAlaLeuPheLysH 2462
QY      914 AGAGCCCGAGAACGATCAGGGGTTGTGTCCAGGCTGGAAAGATGACAGAGAGTA 973
Db      2462 isLysGluGlnLeuArgAlaGluIleLeuLysLysArgAlaLeuLeuAspLysAspLeuG 2482
QY      974 AGTTGCGAG-----GGAAGGGCGCTCAGACATCAAGATGAGGAGCTGGAGACTTG- 1026
Db      2482 InileGluValGlnGluGluLeuLysArgAspLeuLysLysLysGluLysAspLeuM 2502
QY      1027 -----GCGCGAC 1033
Db      2502 etGlnLeuAlaGlnAlaThrAlaValAlaAlaProCysProValThrProAlaProp 2522
QY      1034 CGAGCTGAATGTAGGGTTACGACCC----- 1062
Db      2522 roAlaProAlaProProProProProProProProProProProProProProProPro 2542
QY      1062 ----- 1062
Db      2542 euLeuSerThrProThrLeuProAlaAlaSerGlnLysArgLysArgGluGluLysA 2562
QY      1062 ----- 1062
Db      2562 spSerSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2582
QY      1063 -----AACGCCCTGATTGCTTGGCGGAGCTCACAAC---AACAGGTTTATGATT 1114
Db      2582 ysAspThrLysLeuTyCysIleCysLysLysLysLysLysLysLysLysLysLysLys 2602
QY      1115 GCTGTACCGCTGTGAAGAAATGTTTCATCGCGATTTGTGGCATTTCTGAGGCTCGAG 1174
Db      2602 lyCysAspArgCysGlnAsnTrpTyHisGlyArgCysValGlyIleLeuGlnSerGluA 2622
QY      1175 GGAGGCTTTTGAAGAAAGTGGGAGAGCTATATCTGCCCAACTGC 1221
Db      2622 laGluleuIle-----AspLutyrValCysProGlnCys 2633

```

RESULT 9

```

Q818F0 PRELIMINARY; PRT; 542 AA.
AC Q818F0;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
OS Chromosome scaffold protein p85.
OC Euplotes crassus.
OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
OC Euplotida; Euplotidae; Moneuplotes.
OX NCBI_TaxID=5936;
RN [1]
RP SEQUENCE FROM N.A.
RA Sharp S.I., Pickrell J.K., Jahn C.L.;
RT "The identification of a novel chromosome scaffold protein that
RT associates with Tec elements undergoing en masse elimination in
RT Euplotes crassus."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY154577; AAK75020.1; -.
SQ SEQUENCE 542 AA; 60264 MW; F0A83B5E3AA771A1 CRC64;

```

```

Alignment Scores:
Pred. No.: 0.000101 Length: 542
Score: 189.00 Matches: 113
Percent Similarity: 35.55% Conservative: 101
Best Local Similarity: 18.77% Mismatches: 204
Query Match: 4.03% Indels: 184
DB: 5 Gaps: 24

```

US-09-787-016A-1 (1-2610) x Q818F0 (1-542)

```

QY      211 GTTGTATTATTAGGATTTTCAGGGAAGATGTCTCAAGCTTTCACTGTGTGGACGAGTATGAC 270
Db      6 IleAlaThrLysValAlaGlyLysGlyLeuLysThr-----LysGlyLysLys 21
QY      271 GACAAAGCGAGCCCGAGCAATGAGGAGCCACTAAGGCCATCAAAACCCACC----- 321
Db      22 ThrLysAlaAlaGluLysSerLysLysLysLysLysLysLysLysLysLysLysLys 41
QY      322 -----AGCAAAAGAGTTTCAGGAAACATCATGGGGTTTTCGAAGG 357
Db      42 AsnProSerAspGluGluAspIleAsnLysGluSerAlaArgThrSerHisThrAsnLys 61
QY      358 ACCACTATCGCCAGAGAGGCGGAGGAGCGGAGGAGCGGAGGCTGACCCACTGGAGCGGCCA 417
Db      62 SerGluLysSerAspValGluGluAsnAsnAspHisAspAla----- 75
QY      418 CCCCACAGCAGCAGCTGGGCTGTCCCTGGCGGCGAGTGGGAGGAGCAG----- 465
Db      76 ---ProLysGlnGluValLysLysThrArgLysThrSerSerArgLysThrLysAlaPro 94
QY      466 CCCAAGCGCACTGAGCGGCTGAGCAGATTCCTGACCATTCGCGGCGCGCGGAGGAGG 525
Db      95 ProLysLysThrThrLysGluAsnLys-----ValSerGlyLysLysArgThrArgLys 112
QY      526 AGCATGCTGTCTCCCTGAGGATTTCTGCTGAGCCACGCTCTGCGCCGCCACACAGCC 585
Db      113 ValSerLysLysGluPheAspAspGlu----- 121
QY      586 GAGCAGGCTCCGAGGAGGCGCTGGAAGCGCTTCTGAG-----ACCAGAAGCGGCC 639
Db      122 GluValSerGluAspLysAsnValAlaLysSerSerSerIleLeuArgArgSerArg 141
QY      640 CAGTCTGCTTCCACAGCTGTGAAGAACGACGAGCTCTTCTGAAAGGTGAAGAGGAGG 699
Db      142 LeuAlaAlaAlaAsnAlaAsnLysLysLysLysLysLysLysLysLysLysLysLys 156
QY      700 GATGACCACTGACACCTCCGATAGTCACAGGATGCGCTGACCTTGAAGAGCTTCAG 759
Db      157 ---GluAsnGlnAsnSerSerAspGluGluValLysLysLysLysLysLysLysLys 172
QY      760 AATCCCTTCCAGAGGCGGAGCAGGAGCCCTGAGAGGCGCTCAAGAGGATCCAG 819
Db      173 ---LysSerLysArgGlnAlaLysLysLysLysLysLysLysLysLysLysLysLys 189
QY      820 AGTCGCTCGGAGAGAGCGCGGAGGAGGCTCCCGCGAGACTGTGGGCTCCAGGCC 879
Db      190 AlaGluGluGluGluSerGluLysLysLysLysLysLysLysLysLysLysLysLys 209
QY      880 AGTGACACTGTGGAGGCGTCTCTG---CCGAGTAGCAGAGGAGCCCGAGNACGATCAGGG 936
Db      210 ThrLysProAlaLysLysValLysProThrLys----- 221
QY      937 GTTGTGTCGCGAGCTGGGAAAGATGACAGAGAGATAAGTTGGAGGAAAGCGGCTCAG 996
Db      222 -----AlaAlaLysThrAlaLysLysLysLysLysLysLysLysLysLysLys 232
QY      997 GACATCAAGATGAGGAGCTCGAGACTTGGCCCGACCGAAGCCCTGAATGTGAGGTTAC 1056
Db      233 ---GluAspAspGluProSerGluAspSerGluSerGlnGlyGlyGlyGlyGlu 250
QY      1057 GACCCCAACGCTCTATTGATTTGCGCGAGCTCAACACAGGTTTATGATGTGC 1116
Db      251 GlnGluGluAla-----LysSerGlnAspSerGlnLys----- 261
QY      1117 TGTGACCGCTGTGAAGATGTTTCATGGCGATTTGTGGGCAATTTCTGAGGCTCGAGG 1176
Db      262 -----AspSerGlyGlyGlyGlyGly 268
QY      1177 AGGCTTTTGGAAAGCAATGGGAGAGACTATATCTGCCCAACTGCACCATTTCTGCAAGTG 1236
Db      269 ---GluGluAsnAspGlu-----GluMet 275

```


QY 994 CAGGACATCAAGATGAGGAGCTGGAGACTTGGCCGAGCCAGCCTGAATGTGAGGT 1053
Db : : : : :
QY 275 AlaLysLeuAlaProAlaLysLysGlyAlaSerSerSerSerSerSerSerSerSerSer 294
QY 1054 TAGACACCCCAAGCCCTGATTGATTCATTCGCGCCAGCCTCACACAACAGGTTTATGATT 1113
Db : : : : :
QY 295 GluAlaProLysLysAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 306
QY 1114 TCGTGTGACCGCTGTGAGAGATGGTTTCATGCGGATTTGCGGATTTCTGAGCTCGA 1173
Db : : : : :
QY 307 : : : : : 310
QY 1174 GCGAGGCTTTGGAAGGAATGGGAGAGACTATATCTGCCCCAACTGCCATTCTGCAA 1233
Db : : : : :
QY 311 : : : : : 314
QY 1234 GTGAGATGAGACTCATTACGAAACGCGAGATCAGCAGGAGCTTAATGAGAGACTGGA 1293
Db : : : : :
QY 315 LysAlaAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 334
QY 1294 GATGCTGATGGCAGCCTGATGACACTAGGACATAGCAGAGTCTAGCGAAGAC 1353
Db : : : : :
QY 335 ProAlaLysAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 354
QY 1354 CAAGGATTAAGGAGTGAATGAGAAAGCTGCAATCCAAAGTGGCAGAGAACTCAAG 1413
Db : : : : :
QY 355 Ser 374
QY 1414 ATCTTCACCGCTGTAGAGCGCTGCTGCTCAAAATGATTTGGCCCCGGTCTGT 1473
Db : : : : :
QY 375 : : : : : 383
QY 1474 CAGTGGCGCAGCCGACTGGTGTACTGACGTAATGACTGTATCTCAACACGCGCA 1533
Db : : : : :
QY 384 : : : : : 386
QY 1534 GCGACAATGAAGTTCTTAAGCTCAGTAAGACAGCAGCAAGCCTTAAGAAAGATG 1593
Db : : : : :
QY 387 : : : : : 400
QY 1594 AAGATGAGCCAGAGAGCCAGCTTCCGAAATGGGTCTCAGCAGGATTAATAATC 1653
Db : : : : :
QY 401 LysAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 417
QY 1654 TCTTCTGTGACAGAGACAGCTCAGCAAAA- 1698
Db : : : : :
QY 418 AspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 437
QY 1699 GTGAG- 1704
Db : : : : :
QY 438 AlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 457
QY 1705 : : : : : 1734
QY 458 LysProAlaLysAlaValAlaLysAlaLysAlaLysSerSerSerSerSerSerSer 477
QY 1735 : : : : : 1764
QY 478 AspGluLysProAlaLysAlaLysLysLysLysLysLysLysLysLysLysLysLys 497
QY 1765 AGCAGCGCGTGTGGCGAGCAGATCAATTAATGCAATGCAAGGCA- 1818
Db : : : : :
QY 498 SerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 517
QY 1819 ACTGCTGCTCCCTCG 1833
Db : : : : :
QY 518 ThrAlaAlaProAla 522

RESULT 11
Q95YM2
ID Q95YM2
AC Q95YM2;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE I-connectin.
GN I-CON.
OS Procamburus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacoidae; Cambaridae; Procamburus.
OX NCBI_TaxID=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423462; PubMed=11532946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RL sarcomeres of crayfish claw muscle.";
CC EMBL J. 20:4826-4835(2001). SH3 DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB055861; BAB64297.1; -.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000577; FGGY kin.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGG2; 13.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; 3.
DR PROSITE; PS50835; IG_LIKE; 49.
DR PROSITE; PS50002; SH3; 1.
KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Alignment Scores:
Pred. No.: 0.000235 Length: 17352
Score: 188.50 Matches: 140
Percent Similarity: 34.11% Conservative: 80
Best Local Similarity: 21.71% Mismatches: 201
Query Match: 4.02% Indels: 224
DB: 5 Gaps: 30

US-09-787-016A-1 (1-2610) x Q95YM2 (1-17352)
QY 268 GACGACAAAGGCGACCCGAGCAATGAGAG-----GCACCTAAGCCATCAACCC 318
Db 6208 AspGluGluLysLeuProSerAspGluGlnProLysLeuLysLysLysAlaValLysPro 6227
QY 319 ACCACAAAGAGTTTCAGGAACATGGGGTTTTCGAGGACCCATCTATGCCAAGCGAGAG 378
Db 6228 LysLysAspGluGluGluGluLeuProSerTrpArgGlyLysArgLeuProLysGlu 6247
QY 379 GCGCGAGGGGAC-----GCGAGGCTGACCCACTGAGCGG--- 414
Db 6248 AspGluLysGluGluLeuThrLeuLysProPheLysLysValLysProLysGluProLys 6267
QY 415 CCACCCCCACACACACAGCTGGCTGTCCTCGCGCGAGTGGGAGGCGAGCCCAAGCC 474
Db 6268 ProSerProLysLeuLysProGlyLysProTyrGluProGluLeuProGluProGluLys 6287
QY 475 ACTGAGCGCGTGGAGCAGTTCTGACCATTCGCGCGCGCGCGCGAGGAGCGATGCT 534
Db 6288 Ser-----Pro 6289
QY 535 GTCTCCTCGAGGATTCTGTGAGCCACGCTCTGCCCGCCACAGACGCCGAGACAGCC 594
Db 6290 Leu-----GluProTyrSerLysProGluLysGlu-----Val 6300
QY 595 TCCGAGGGCAGCGTGCAGAAAGCGCTTCTGAGACCAAGAGCGGCCCGCCAGTCTCTCCACA 654

```

Db 6301 SerAspLysIleProGluLeuAlaLysGluValSerArgLysProGluGlu----- 6317
QY 655 GCTGTGAGGAAACGACGACGCTCTTCTGAAAGAGTGAAGAGGAGGATCACCACGATGAC 714
Db 6318 -----ProGluLysProGluGluProGluLysLeu-----AspGluSerAsp 6331
QY 715 ACCTCGATAGTACAGCGATGCGCTGACCTTGAAGAGCTTCAAGATCGCTTCGACGG 774
Db 6332 LysLysProAspGluSerGluThrIleThrGluLysPro-----LysArgArgLeu 6348
QY 775 AACCGGGAACAGGAGCCACTGAGAGCCCTGAAAGGGATCCAGAGTCGCTGCGGAAG 834
Db 6349 GlnLysLeuLysAlaProThrGluGluLysPheGluIleProLysValThrLeuArgLys 6368
QY 835 AACGCGCGG-----GAGAGGGTCCCGCCGAGACTGTGGCTCCGAGGCC 879
Db 6369 ThrSerGlnLysValPheValProGluGluValThrLeuGluThrValGluLeuGluHis 6388
QY 880 AGTGACACTGTGAGGGCGTCCGCCCAGT----- 909
Db 6389 ValGluThrProGluValValGluProGluValGluLysArgValTrpSerProProPro 6408
QY 909 ----- 909
Db 6409 GluTyrGluThrTyrValProGluGluIleProGluLysGluProValGluLeuGluLys 6428
QY 910 -----AACGAGGCCCCGAGACCATCAGGCGGTT 939
Db 6429 TyrGluLysTyrGluProProThrLysProLysAspGluGluGluLysGlyLys 6448
QY 940 GTGTCCAGGCTGGGAAGAT-----GACAGAGAGAGTAAAGTTG--- 978
Db 6449 TyrGluArgLysProLysAspLysProGluProGluGluAspArgLysLeuLysLeuGly 6468
QY 979 GAGGGAAGCGGCTCAGGACATCAAGATGAGAGCGCTGGAGC----- 1023
Db 6469 LysGlyLys-----LeuArgProGluGluGluGluGluGluLysLysLys 6484
QY 1024 -----TTGGGCGGACCGAAG---CTGATGTGAGGTTAOCACCCCAAGCGCTGTAT 1074
Db 6485 ProProLysArgArgProLysSerProGluLysGluAlaGluLysProGlnLeuLysPro 6504
QY 1075 TGCATTGCGCGGACGCTCAACAACAGGTTTATGATTGTGTGACGCGCTGGAAGAA 1134
Db 6505 IleProLysLysLysProGlu----- 6511
QY 1135 TGGTTTCATGGCGATGTGTGGGCAATTCGAGCTCGAGGAGGCTTTTGGAAAGGAAT 1194
Db 6512 -----GluGluLysLysAspLysValThrProLysPro 6522
QY 1195 GGGGAAGACTATATCTGCCCAAACTGCCCAAT-----CTGCAA 1233
Db 6523 GlyLysLys-----ProSerLysLysIleProAspArgGluProValGluLeuGlu 6539
QY 1234 GTCAGATAGACTATCTTCCGAAACCGGAGTACGAGCAAGCT----- 1278
Db 6540 ProPheGluArgThrGluProGluIleLeuAspLysAspLysValProLeuGluLysPro 6559
QY 1279 -----AAATGGAGACTTTCGAGATGCTGTGATGTCACCGATTTGACAGTATAGCAACA 1329
Db 6560 LeuLysProLysProGluProLysGluLysProSerIleGluProGluAlaProLysPro 6579
QY 1330 ATAGACAGAAGTCTAGCGAAGCAACAGGATTAAGGGTTAGATTGAAGAGCTGCAAT 1389
Db 6580 LeuGluLysProSerGluGluGluGluProLysLysLysLysLysGluArgPro--LysP 6599
QY 1390 CCAAGTGGCAAGAGAACTCAAGATCTTCCAGCTGTGATAGAGGCGCTGGCTCA 1449
Db 6599 roGluLysLysGluGluGluAlaGluValProser----- 6610
QY 1450 AAATGTATTGCGCCCGGGTCTGTCTACGTGGCGCAGCCGACTCGGTGTACTCCAGTAAT 1509

```

```

Db 6611 -----Trp-----ArgGlyLysArg----- 6615
QY 1510 GACTGTATTCCTCAACACAGCCCGCAGCAATGAAGTTTCTTAAGCTCAGGTAAGAACAG 1569
Db 6616 --LeuProProLysGluGluGluLysGluIleVal--LeuLysProPheLysLysGlu 6634
QY 1570 AAGCCA---AAGCCTAAAGAAAGATGAAGATGAAGCCAGAGAGCCAGTCTTCCGAAA 1626
Db 6635 LysProGluGluProLysProLysProLysProLysProGlyLysProTyrGluProGlu 6654
QY 1627 TGC----- 1629
Db 6655 IleProGluProGluLysThrProLeuGluProTyrThrLysProAspLysGluLysVal 6674
QY 1630 -----GGTGCTCAGCAGGATTAATAATCTTCTGTGCAAGAGACCACTCCAGAA 1683
Db 6675 ProAspGlyValThrGluProValLysProGluAsp---GluGluLysProLysProGlu 6693
QY 1684 AAAAAAGAGACCACTGCAAGAGCAGTGGTGGTCCCTGCGCGAGTGAAGCACTCGGG 1743
Db 6694 ---GluGluIleLysProLysLysGluArgIleLysProAspLysAspGlu----- 6709
QY 1744 AAGGAGCGCTGTGAGAGCAGCAGCGCTGCTGG----- 1779
Db 6710 -----GluValGluThrProSerTrpArgGlyLysArgLeuProProLys 6724
QY 1780 GCGAGCGATCAATTAACAATGCACTAAAGCCA-----GAAAAGACTGCTGCTCCC 1830
Db 6725 GluGluAspLysGluGluIleThrLeuLysProPheLysLysGluLysProLysGluPro 6744
QY 1831 TCGCGGTCA 1839
Db 6745 LysProSer 6747

RESULT 12
Q8C969 PRELIMINARY; PRT; 443 AA.
ID Q8C969; AC Q8C969;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN A630082K20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK042834; BAC31377.1; --
DR MGD; MGI:2443388; A630082K20RIK.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003347; TF_JmjC.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF02373; JmjC; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00558; JmjC; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS01359; ZF_PHD 1; 1.
DR PROSITE; PS0016; ZF_PHD_2; 1.
DR Hypothetical protein.
KW NON_TER 443
SQ SEQUENCE 443 AA; 49944 MW; C543F487717C4682 CRC64;

```

Alignment Scores: 0.000134 Length: 443
Pred. No.: 187.00 Matches: 78
Score: 187.00 Conservatives: 44
Percent Similarity: 34.46% Mismatches: 118
Best Local Similarity: 22.03% Indels: 114
Query Match: 3.99% Gaps: 18
DB: 11

US-09-787-016A-1 (1-2610) x Q8C969 (1-443)

QY 982 GGAAAGCGCGCTCAGGACATCAAGATGAGGAGCTGGAGACTTGGCGCGACCGAAGCCT 1041
Db 15 GlyAlaAlaAlaAlaGlySerValSerAlaProGlyArgAlaSerAlaProPro 34
QY 1042 GAATGTGAGGTTACAGCCCAAGCCCTGATTGCTATGCTGCGCAGCTCACAAC--- 1098
Db 35 -----ProProValTyrCysValCysArgGlnProTyrAspVal 48
QY 1099 AACAGGTTTATGATTGCTGTGACCGCTGTGAGATGGTTTCATGGCATTGTGGGCG 1158
Db 49 AsnArgPheMetIleGluCysAspValCysArgGlnProTyrAspVal 68
QY 1159 ATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAC 1218
Db 69 ValGluGluHisAlaValAspIleAspLeu-----TyrHisCysProAsp 84
QY 1219 TGACACATCTCTCAAGTCAGGATGAGACTCATTGAGACGSCGATCAGCAGGAAGCT 1278
Db 85 CysAlaAlaLeu-----HisGlySerLeuMetIleGlySerArgArg 98
QY 1279 AATGTGAGACTCGGAGAT-----GCTGATGGCAGCGATTGTACAAAGTATAGAA 1326
Db 99 AsnTyrHisArgHisAspTyrThrGluValAspGlySerIleProValGlnAlaGly 118
QY 1327 ACA----- 1329
Db 119 ThrArgAlaPheValIysGluLeuArgSerArgValPheProSerAlaAspGluIleIle 138
QY 1330 -----ATAGACAGAACTCTAGCGAAGCAACCAAGG-----ATA 1362
Db 139 ValIysMetHisGlySerGlnLeuThrGlnArgTyrLeuGluIleHisGlyPheAspVal 158
QY 1363 AAGGTAGATTGAGAAAGCTGCAATCCAAAGTGGCAGAGAACTCAAGATCTCCAG 1422
Db 159 ProIleMetValProIleValLeuAspAspLeuGly-----LeuArgLeuProSer 174
QY 1423 CQTGTGATAGAGCGCTGCTGCTCAAAATGATTGGCCCGCGTGTCTCAGTGGCG 1482
Db 175 ProAlaPheSerValMetAspValGluArgTyrValGly----- 187
QY 1483 CAGCCGAGCTGGGTACTGCAAGTATGCTATCTCTCAACAGCGCGCAGCAATG 1542
Db 188 ---GlyAspIleValIleAspValIleAsp---ValAlaArgGlnAlaAspSerIleMet 205
QY 1543 AAGTTTCTAGCTCAGGTAAAGACAGACAGCAAGCCTTAAGAAAGATGAAGATGAAG 1602
Db 206 ThrLeuHisAsn-----TyrValIleTyrPheMetAsn 216
QY 1603 CCAGAGAAGCCAGCTCTTCCGAAATGCGGTGCTCAGGACGATTAATAATCTCTCTGTG 1662
Db 217 ProAspArgProIleVal-----LeuAsnValIleSerLeu 228
QY 1663 CACAGAGACCGCTCCAGAAAAAAGACAGACACAGTGAAGAGCGGAGTGGTCCCT 1722
Db 229 -----GluPheSerAspThrIleMetSerGluLeuValGluValPro 242
QY 1723 GCGCGAGTGAAGCACTCGGAGGAAGACGCTTGTGAGAGCAGCAGCGCTCTGGGCG 1782
Db 243 AspIleAlaArgIleLeu-----SerTyrVal 251
QY 1783 AGCGATCAATTAATGCAATGAGTAAAGCAGAAAGACTGCTGTCTCCCTCGCTCACTG 1842
Db 1783 AGCGATCAATTAATGCAATGAGTAAAGCAGAAAGACTGCTGTCTCCCTCGCTCACTG 1842

252 -----GluAsnTyr-----TyrProAspAspSerValPheProIysProPheVal 266
QY 1843 TTGTATAAATGATG-----TATCAGCTAGGCTTGGC 1875
Db 267 GlnIysTyrCysLeuMetGlyValGlnAspSerTyrThrAspPheHisIleAspPheGly 286
QY 1876 CTCCTGAGGAGCCCTCCGCTTCTTCTGGATAGCCATCCCTCG 1917
Db 287 -----GlyThrSerValTyrTyrHisValLeuTyr 296

RESULT 13
Q8C9E0 PRELIMINARY; PRT; 473 AA.
AC Q8C9E0;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN A630082K20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 50,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK042327; BAC1226.1; -
DR MGD; MGI:2443388; A630082K20RIK.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR003347; TF_JmjC.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF02373; JmjC; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00558; JmjC; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01186; EGF 2; 1.
DR PROSITE; PS01359; ZF_PHD 1; 1.
DR PROSITE; PS00016; ZF_PHD 2; 1.
KW Hypothetical protein.
FT NON TER 473 473
SQ SEQUENCE 473 AA; 53409 MW; 94C94378609C7039 CRC64;

Alignment Scores:
Pred. No.: 0.000136 Length: 473
Score: 187.00 Matches: 78
Percent Similarity: 34.46% Conservatives: 44
Best Local Similarity: 22.03% Mismatches: 118
Query Match: 3.99% Indels: 114
DB: 11 Gaps: 18

US-09-787-016A-1 (1-2610) x Q8C9E0 (1-473)

QY 982 GGAAAGCGCGCTCAGGACATCAAGATGAGGAGCTGGAGACTTGGCGCGACCGAAGCCT 1041
Db 15 GlyAlaAlaAlaAlaGlySerValSerAlaProGlyArgAlaSerAlaProPro 34
QY 1042 GAATGTGAGGTTACAGCCCAAGCCCTGATTGCTATGCTGCGCAGCTCACAAC--- 1098
Db 35 -----ProProValTyrCysValCysArgGlnProTyrAspVal 48
QY 1099 AACAGGTTTATGATTGCTGTGACCGCTGTGAGATGGTTTCATGGCATTGTGGGCG 1158
Db 49 AsnArgPheMetIleGluCysAspValCysArgGlnProTyrAspVal 68
QY 1159 ATTTCTGAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAAC 1218
Db 69 ValGluGluHisAlaValAspIleAspLeu-----TyrHisCysProAsp 84
QY 1219 TGACACATCTCTCAAGTCAGGATGAGACTCATTGAGACGSCGATCAGCAGGAAGCT 1278
Db 85 CysAlaAlaLeu-----HisGlySerLeuMetIleGlySerArgArg 98
QY 1279 AATGTGAGACTCGGAGAT-----GCTGATGGCAGCGATTGTACAAAGTATAGAA 1326
Db 99 AsnTyrHisArgHisAspTyrThrGluValAspGlySerIleProValGlnAlaGly 118
QY 1327 ACA----- 1329
Db 119 ThrArgAlaPheValIysGluLeuArgSerArgValPheProSerAlaAspGluIleIle 138
QY 1330 -----ATAGACAGAACTCTAGCGAAGCAACCAAGG-----ATA 1362
Db 139 ValIysMetHisGlySerGlnLeuThrGlnArgTyrLeuGluIleHisGlyPheAspVal 158
QY 1363 AAGGTAGATTGAGAAAGCTGCAATCCAAAGTGGCAGAGAACTCAAGATCTCCAG 1422
Db 159 ProIleMetValProIleValLeuAspAspLeuGly-----LeuArgLeuProSer 174
QY 1423 CQTGTGATAGAGCGCTGCTGCTCAAAATGATTGGCCCGCGTGTCTCAGTGGCG 1482
Db 175 ProAlaPheSerValMetAspValGluArgTyrValGly----- 187
QY 1483 CAGCCGAGCTGGGTACTGCAAGTATGCTATCTCTCAACAGCGCGCAGCAATG 1542
Db 188 ---GlyAspIleValIleAspValIleAsp---ValAlaArgGlnAlaAspSerIleMet 205
QY 1543 AAGTTTCTAGCTCAGGTAAAGACAGACAGCAAGCCTTAAGAAAGATGAAGATGAAG 1602
Db 206 ThrLeuHisAsn-----TyrValIleTyrPheMetAsn 216
QY 1603 CCAGAGAAGCCAGCTCTTCCGAAATGCGGTGCTCAGGACGATTAATAATCTCTCTGTG 1662
Db 217 ProAspArgProIleVal-----LeuAsnValIleSerLeu 228
QY 1663 CACAGAGACCGCTCCAGAAAAAAGACAGACACAGTGAAGAGCGGAGTGGTCCCT 1722
Db 229 -----GluPheSerAspThrIleMetSerGluLeuValGluValPro 242
QY 1723 GCGCGAGTGAAGCACTCGGAGGAAGACGCTTGTGAGAGCAGCAGCGCTCTGGGCG 1782
Db 243 AspIleAlaArgIleLeu-----SerTyrVal 251
QY 1783 AGCGATCAATTAATGCAATGAGTAAAGCAGAAAGACTGCTGTCTCCCTCGCTCACTG 1842
Db 1783 AGCGATCAATTAATGCAATGAGTAAAGCAGAAAGACTGCTGTCTCCCTCGCTCACTG 1842


```

Db 69 ValGluGluHisAlaValAspLeu-----TyrHisCysProAsp 84
QY 1219 TGCAACCATCTCGAAGTGCAGGATGACATTCATTAGAAACCGCAGCATCAGCAGGAAAGCT 1278
Db 85 CysAlaAlaLeu-----HisGlySerSerLeuMetLysLysArg 98
QY 1279 AAATGAGACCTGGAGAT-----GCTGATGGCACCCGATTGTACAGTATAGCA 1326
Db 99 AsnTrpHisArgHisAspTyrThrGluValAspAspGlySerLysProValGlnAlaGly 118
QY 1327 ACA----- 1329
Db 119 ThrArgAlaPheValLysGluLeuArgSerArgValPheProSerAlaAspGluIleIle 138
QY 1330 -----ATAGACGAGAGCTTAGCGAAGACCAAGGG-----ATA 1362
Db 139 ValLysMetHisGlySerGlnLeuThrGlnArgTyrLeuGluLysHisGlyPheAspVal 158
QY 1363 AAGGTTAGATTGAGAAAGCTGCAATCCAAAGTGGCAAGAGAAATCAAGATCTTCCAG 1422
Db 159 ProIleMetValProLysLeuAspAspLeuGly-----LeuArgLeuProSer 174
QY 1423 CCTGTGATAGAGCGCTGTGTGCTCAAAATGATTTGGCCCGCGTGTCTACGTGGCG 1482
Db 175 ProAlaPheSerValMetAspValGluArgTyrValGly----- 187
QY 1483 CAGCCGACCTGCTGTACTCTGAGTAATGACTGTATCTCTCAACACGCGCAGCGACAATG 1542
Db 188 ---GlyAspLysValIleAspValIleAsp---ValAlaArgGlnAlaAspSerLysMet 205
QY 1543 AAGTTTCTAAGCTCAGGTAAAGAACAGACGACCAAGCCTAAAGAAAGATGAAGATGAAG 1602
Db 206 ThrLeuHisAsn-----TyrValLysTyrPheMetAsn 216
QY 1603 CCAGAGAACCCAGCTCTTCGAAATCGGTGCTCAGCAGGATTAATAATCTCTCTGTG 1662
Db 217 ProAspArgProLysVal-----LeuAsnValIleSerLeu 228
QY 1663 CACAGAGACCGCTCCAGAAAAAAGACAGACACAGTGAAGAGGCGAGTGTGTCCCT 1722
Db 229 -----GluPheSerAspThrLysMetSerGlnLeuValGluValPro 242
QY 1723 GCGCGAGTGAAGCATCTCGGAGGAAGACAGCTTGTGTGAGACGACGCGCGTGTGGCG 1782
Db 243 AspIleAlaArgLysLeu-----SerTrpVal 251
QY 1783 AGCGATCACAAATACAAATGAGTAAGCCAGAAAGACTGCTGCTCGCGCTCACTG 1842
Db 252 -----GluAsnTyr-----TrpProAspAspSerValPheProLysProPheVal 266
QY 1843 TTGTATAAATCTATG-----TATCACTAGGGGTGGC 1875
Db 267 GlnLysTyrCysLeuMetGlyValGlnAspSerTyrThrAspPheHisIleAspPheGly 286
QY 1876 CTCCTGGACCCCTCCCTCTTCTTCTGATAGCCATCCCTGG 1917
Db 287 -----GlyThrSerValTrpTyrHisValLeuTrp 296

```

RESULT 14

```

Q7SZE4 PRELIMINARY; PRT; 563 AA.
AC Q7SZE4;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
[1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN-AB; TISSUE=Body;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altshuler S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards R.F., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalski U., Smalios D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB; TISSUE=Body;
RA Klausner R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052918; AAH52918.1; --
KW Hypothetical protein.
SQ SEQUENCE 563 AA; 66587 MW; A0834B9C49192E83 CRC64;

```

Alignment Scores:

```

Pred. No.: 0.000179 Length: 563
Score: 185.50 Matches: 90
Percent Similarity: 35.64% Conservative: 49
Best Local Similarity: 23.08% Mismatches: 90
Query Match: 3.96% Indels: 161
DB: 17 Gaps: 17

```

US-09-787-016A-1 (1-2610) x Q7SZE4 (1-563)

```

QY 1006 GATGAGGACCTGGA---GACTTGGCGCCGACCGACCTGAATGTGAGGTTACGACCCC 1062
Db 9 AspGlnThrProGlyLeuAspAsnSerMetGluLysGlyGlu----- 22
QY 1063 AACGCC---CTGTATTGCAATTTCCGCCAGCTCCACAAACAGAGTTTATGATTGCTGT 1119
Db 23 AsnAlaProLeuTyrCysIleCysArgLysSerAspIleAsnCysPheMetIleGlyCys 42
QY 1120 GACCGCTGTGAGAATGGTTTCATGCGGATTGTGCGCATTTCTGAG----- 1167
Db 43 AspAsnCysAsnGluTrpPheHisGlyHisCysIleAsnValThrGluLysMetAlaLys 62
QY 1167 ----- 1167
Db 63 AlaIleArgGluTrpTyrCysGlnGlnCysArgAlaArgAspProSerLeuSerIleArg 82
QY 1167 ----- 1167
Db 83 TyrArgLysLysAsnArgAspLysAspValGluProGluArgValGluLysArgSerSer 102
QY 1168 -----GCTCGAGGAGGAGCTTTTGGAAAGAGATGG----- 1197
Db 103 ThrProGluTyrLysIleAspLysArgArgGlySerLysValLysArgSerAlaArgMet 122
QY 1198 -----GAAGACTATATCTGCGCAAACTGCACCAT 1227
Db 123 CysGlyGluCysGluProCysThrArgThrGluAsp-----CysGlyHisCys-AspPhe 140
QY 1228 CTCG-----AAGTCAGGATGAG 1245
Db 140 eCysLysAspMetLysLysPheGlyGlyProAsnLysIleArgGlnLysCysArgLeuArg 160

```

```

QY 1245 A-----CTCATTCAGAAACCGCAGATCAGCAGGAAGCTTAATGGAGACCTGGAGATG-- 1297
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
160 gGlnCysValValArgAlaArgLysMetLeuArgValArgAspGluGluPheSerLeuArg 180
QY 1298 -----CTGATGCACCGA----- 1310
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
180 gGluArgLysAspAsnMet-HisArgAspArgGlyTyrSerAspAspTyrAspGluA 200
QY 1311 -----TTGTCAAGTATAGGAACAATAGA----- 1334
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
200 snAspMetAspLeuTyrGluHisTyrLysAspArgAsnAlaSerTrpGlySerGluAspA 220
QY 1335 -----GCAGAGTCT 1344
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
220 spAspGlyGlnLeuTyrSerProValProArgLysLysAlaLeuLysValLysVal 240
QY 1345 AGCAGACACCAAGGATTAAGGGTAGAATTGAGAAAGCTGCAATCAAGTGCACAGAG 1404
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
240 ysArgArgAspLysLysPheAspLysLysGluSerArgArgHisLysGlnLysGlnL 260
QY 1405 AAATCAAGAT---CTTCCAGCTGTGATAGA-----GGCGCTGTGTGCC 1446
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
260 ysHisArgAspArgLeuArgHisSerAspArgThrAspGlyArgHisGlyGly-AspThr 279
QY 1447 TCAAAATGTTATGGCCCGGCTGTGTACGTGCGCAGCCGACTCGGTGTACTGCAGT 1506
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
280 GlnGlnCysLeuGlyProAsnCysIleGluProAlaArgProAsnSerLysTyrCysSer 299
QY 1507 AATGACTGTATCTCAACACCGCGCAGCGCAGCAATGAATGTTCTTAAGCTCAGGTAAGAA 1566
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
300 GluAspCysGlyMetLysLeuAlaAlaAsnArgile-----TyrGlu 313
QY 1567 CAGAGCCAAAGCTTAAGAAAGATGAAGATGAAGCCAGAGACCCAGTCTTCCGAAA 1626
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
314 ValLeuProGlnArgIleGlnGlnTrpGlnGlnSerPro----- 326
QY 1627 TGGCGTGCT-----CAGGCGAGTATTAAATCTTCTGTGCAAGAGACAGCTCCA 1680
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
327 CysIleAlaGluGlnGlnGlyLysLysGlnLeuLysArgGlnGlnGlnAl 346
QY 1681 GAAAAAAGAGA---CCACAGTGAAGAGCAGTGTGTCTGCGCGAGTGAAGCA 1737
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
346 alaArgMetArgLeuAlaGluMetGluArgArgPheHisGluLeuGluGlyIleIleAl 366
QY 1738 CTCGGAAGGAGCAGCTGTG 1759
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
366 aLysAlaLysGlnLeuVal 373

RESULT 15
O74508 PRELIMINARY; PRT; 424 AA.
AC O74508;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SPCC594.05c.
GN SPCC594.05c.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h;
RA Rieger M., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031523; CAA20664.1;
DR PIR; T41449; T41449.
DR GeneDB_Spombe; SPCC594.05c;
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

```

```

DR InterPro: IPR001965; Znf_PHD.
DR Pfam: PF00628; PHD; 1.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS01359; ZF_PHD_1; 1.
DR PROSITE: PS0016; ZF_PHD_2; 1.
SQ SEQUENCE 424 AA; 48683 MW; 2963605C3DFC0B9 CRC64;

Alignment Scores:
Pred. No.: 0.000183 Length: 424
Score: 185.00 Matches: 96
Percent Similarity: 35.05% Conservative: 47
Best Local Similarity: 23.53% Mismatches: 148
Query Match: 3.95% Indels: 117
DB: 3 Gaps: 15

US-09-787-016A-1 (1-2610) x O74508 (1-424)
QY 535 GTCTCCCTGGAGGATTCGTGTGAGCCCGCAGCTCTCCCGCCGACACAGAC-----GCCGAG 588
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
20 ValLysPheGluAspSerAsnArgGlyThr-----IleThrAspPheHisIleGlu 36
QY 589 ACAGCTCTCGAGGGCGAGGTGGAAGCGCTTCTGAGACCGAGAGCGGCCCGCCAGTCTGCT 648
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
37 ThrAlaAsnAsnGluGluLysAspAlaAsnValIleLeuAsn----- 51
QY 649 TCCACAGCTGTGAAGAGCAGCAGCAGCTCTTCTGAAAGCTGAAAGGGGATGACACAC 708
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
52 -----LysSerValLysMetGluValGluValAsnGly-----His 64
QY 709 GATGACACTCCGATAGTGCAGCAGCGATGCCCTTGAAAGAGCTTCAGAAAGAGCTTCAGAATCGCCTT 768
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
65 ValAspSerSerSerThrGluThrAsp-----Ile 74
QY 769 CCGAGGAACGGGAGCAGGAGCCACTGAGAGCCCTGAAAGGGATCCAGAGTCCGCTG 828
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
75 GluMetGlnValIleGlnGlnProThrIleProLysLysProValSerAlaHisArg 94
QY 829 CGGAAGAACGGGAGGAGGAGGTCCCGCCGAGACTGTGGGCTCCGAGGCCAGTGCACACT 888
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
95 ArgGlyProArgLysHisArgLysAsnAla----- 104
QY 889 GTGAGGGCGCTCTGCTCCAGTAGCAGGAGCCGAGAGCATCAGGGGGTGTGTCCAG 948
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
105 -----AsnSerGlnLeuAsnLeuSerThr 112
QY 949 GCTGGGAAGATGACAGAGAGTAAGTTGGAGGAAAGCGCGCTCAGGACATCAAGAT 1008
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
113 Ala----- 113
QY 1009 GAGGAGCTGGAGACTTGGGCCGACCGAAGCCTGAATGTGAGGGTTACGACCCCAAGCC 1068
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
114 -----AspHisGlnArgPro----- 118
QY 1069 CTGTATTGATTGGCCGCGCAGCTCAACACAGGTTTATGATTGCTGTGACCGCTGT 1128
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
119 LeuTyrCysIleCysGlnLysProAspAspGlySerTrpMetLeuGlyCysAspGlyCys 138
QY 1129 GAAGAAATGGTTTCATGCGCATTTGTGGCATTTCTGAGGCTCGAGGAGGCTTTTGGAA 1188
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
139 GluAspTrpPheHisGlyThrCysValAsnIleProGluSerTyrAsnAspLeuThrVal 158
QY 1189 AGCAATGGGAGACACTATATCTCCCAAACTGCACCATTTCTGCAAGTGCAGGATGAGACT 1248
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
159 Gln-----TyrPheCysProLysCysThr-----GluGluGly 169
QY 1249 CATTCAAGAAACCGCAGATCAGCAGGAGCTAAATGGAGA-----CCTGGAGAT 1296
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
170 LysGlyIleThrThrTrpLysArgLysCysArgLeuArgGluCysSerAsnProThrArg 189
QY 1297 GCTGATGGCCCGATTGTACAGTATAGGAACAATAGACAGAGTCTACGAGACCAA 1356
Db | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
190 ProAsnSerAsnTyrCys-----SerAspLysHis 199

```

Search completed: April 28, 2004, 11:00:23
Job time : 200.5 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	193	4.1	48.1	488	2	I46014	cylicin II - bovin
2	189.5	4.0	70.6	2	A45990		junctional sarcom
3	185	3.9	42.4	2	T41449		probable phd finger
4	182	3.9	65.9	2	I38073		nucleolar phosphop
5	181	3.9	53.7	2	T13564		microtubule-associ
6	180.5	3.8	66.9	2	JC5562		hepatoma-derived g
7	180	3.8	60.6	2	A43427		neurofilament trip
8	176.5	3.8	111.0	2	I51116		NG-180 - sea lamp
9	174	3.7	73.4	2	B42680		nucleolus-cytoplas
10	170.5	3.6	53.0	2	T21430		hypothetical prote
11	169.5	3.6	99.0	2	I51618		nucleolar phosphop
12	169.5	3.6	168.4	2	FW0957		gravin - human
13	169	3.6	72.9	2	S68191		triadin - human
14	167.5	3.6	108.7	1	OPMSH		neurofilament trip

D_b 454 GluInGluThrArgysGluLysProGlyLysIleSerSerValLeuLysAspLysGlu 473
QY 1759 -----GAGACGACGCGCTGTGGCGAGCGATCACAAATTACAATGCA 1803
D_b 474 LeuThrLysGluLysGluValProAlaSerLeuLysGlySerGluThr 493
QY 1804 GTAAGCCAGAAGACTCTCTCCCTGCCGCTCACTG 1842
D_b 494 LysLysAspGluLysThrSerLysProGluProGlnIle 506

RESULT 3
T41449
probable phd finger transcription regulator - fission yeast (Schizosaccharomyces
C; Species: Schizosaccharomyces pombe
C; Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
C; Accession: T41449
R; Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, September 1998
A; Reference number: Z21994
A; Accession: T41449
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: DNA
A; Residues: 1-424 xRIS>
A; Cross-references: EMBL:AL031523; PIDN:CAA20664.1; GSPDB:GN00068; SPDB:SPCC594
A; Experimental source: strain 972h; cosmid c594
C; Genetics:
A; Gene: SPDB:SPCC594.05c
A; Map position: 3

Alignment Scores:
Pred. No.: 0.000293 Length: 424
Score: 185.00 Matches: 96
Percent Similarity: 35.05% Conservative: 47
Best Local Similarity: 23.53% Mismatches: 148
Query Match: 3.95% Indels: 117
DB: 2 Gaps: 15

US-09-787-016A-1 (1-2610) x T41449 (1-424)
QY 535 GTCTCCTCGAGGATTCGTGAGCCCGCATCTGCTCCCGCCACAGAC-----GCCGAG 588
D_b 20 ValLysPheGluAspSerAsnargGlyThr-----IleThrAspPheHisIleGlu 36
QY 589 ACAGCTCCGAGGGCAGCGTGGAAGCGTTCTGAGACGAGAGCGGCCCGCCAGTCTGCT 648
D_b 37 ThrAlaAsnGluGluLysAspAlaAsnValIleLeuAsn----- 51
QY 649 TCACACAGCTGTGAAGAACGACCAGCGCTCTCTGAAAAGGTGAAAGGGGATGACCAC 708
D_b 52 -----LysSerValLysMetGluValGluGluValAsnGly-----His 64
QY 709 GATGACACTCCGATAGTAGACAGCGATGCGCTGCACCTTGAAAGAGCTTCAGAAATCGCCT 768
D_b 65 ValAspSerSerThrGluThrAsp-----Ile 74
QY 769 CGCAGGAACGGGACACAGAGCCCCTGAGAGCCCTCTGAAAGGGATCCAGAGTGCCTG 828
D_b 75 GluMetGlnValIleGlnProThrIleProLysLysProValSerAlaHisArg 94
QY 829 CGGAAGAAGCGCGGAGAGAGGTCTCCGCCGAGACTGTGTGGCTCCGAGCCAGTGCACCT 888
D_b 95 ArgGlyProArgLysHisArgGlyAsnAla----- 104
QY 889 GTGAGGGGCTCTCCGCCAGTAGACAGAGCCCGAGAACGATCAGSGGGGTGTGTCCAC 948
D_b 105 -----AsnSerGlnLeuAsnLeuSerThr 112
QY 949 GCTGGGAAGATGCACAGAGAGTAGATTGGAGGGAAGGCGCTCAGGACATCAAGAT 1008
D_b 113 Ala----- 113
QY 1009 GAGGACCTTGAGACATTGGGGCCGACCGAAGCTGTGATGTGAGGGTTACGACCCCAAGCC 1068

```

114 -----AspHisGlnArgPro----- 118
1069 CTGTATTGCTATTCGCGCAGCTCACAAACAGGTTTATGATTGCTGTGACGCTGT 1128
119 LeuTyrCysIleCysGlnLysProAspAspGlySerTrpMetLeuGlyCysAspGlyCys 138
1129 GAAGATGTTTCATGCGCATTTGTGGGCTCTGAGGCTCGAGGAGGCTTTTGAA 1188
139 GluAspTrpHeHisGlyThrCysValAsnIleProGluSerTyrAsnAspLeuThrVal 158
1189 AGGAATGGGAGAGACTATATCTGCCAAACTGCACCAATCTCTCAAGTGCAGGATGAGACT 1248
159 Gln-----TyrPheCysProLysCysThr-----GluGluGly 169
1249 CATTCAGAAACCGCAGATCAGCAGGAGCTAAATGGAGA-----CCTGGAGAT 1296
170 LysGlyIleThrThrTrpLysArgLysCysArgLeuArgGluCysSerAsnProThrArg 189
1297 GCTGATGGCAGCATGTTGTACAGTATAGGAACAATAGACAGCAAGTCTAGCGAAGCAAA 1356
190 ProAsnSerAsnTyrCys-----SerAspLysHis 199
1357 GGATAAAGGCTAGAAATTCAGAAA-----GCTGCAAAATCCAGTGGCAGGAAG 1404
200 GlyValAspPheArgGluLysValLysLeuSerThrValGluProSerAlaIleLys 219
1405 AAATCAAGATCTTCAGGCTGTGATAGAGGCGCTGTGCTCAAAATGATTGGCCCC 1464
220 AsnLeuValLeuPheAlaLysLysArgGluLysPheGlnAsnLeuGlyThrValGlyPro 239
1465 GGTGCTGTCACTGCGCGCAGCGGCTGGTGTACTGCGAGTAAGTACTGTATCTCTCAA 1524
240 ThrLeuProSerGlnValProProGluValValTyr-----AsnPheGluIleGluGlu 257
1525 CACGCGCAGCGAATAGTAAAGTTTCTAGCTAGGTAAAGACAGCAGCAAGCCTAAA 1584
258 AlaAsnArgLeuAsnAlaGluIleValGlnLeuAsnLysGluLysGluValAlaSerAsn 277
1585 GAAATGATGATGAGAGCAGCAGCAGCAGCAGCTTCCGAATGCGGTCTC----- 1636
278 LysLysIlePheLeuGluGluIleLysAspSer-SerArgArgAlaValLeuAlaTyrIly 297
1637 ----AGGCGAGTAAATATCTCTTCTGTCACAGACAGCAGCTCCAGAAAAAAGAG 1692
297 sGluArgGluGlyIleLysLysAspLeuLysCysGlyPheAspSerArgLeuLeuPheAsnGly 317
1693 ACCACAGTGAAGAGCGACTGG 1714
317 nGlnGlnMetAsnGluLeuTrp 324

RESULT 4
I38073
nucleolar phosphoprotein p130 - human
C;Species: Homo sapiens (man)
C;Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 24-Sep-1999
C;Accession: I38073; S52292
R;Pai, C.Y.; Chen, H.K.; Sheu, H.L.; Yeh, N.H.
J. Cell Sci. 108, 1911-1920, 1995
A;Title: Cell-cycle-dependent alterations of a highly phosphorylated nucleolar protein p
A;Reference number: I38073; MUID:95386590; PMID:7657714
A;Accession: I38073
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-699 <RES>
A;Cross-references: EMBL:Z34289; NID:9663007; PIDN:CAA84063.1; PID:9663008
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

Alignment Scores:
Pred. No.: 0.000489 Length: 699
Score: 182.00 Matches: 117
Percent Similarity: 36.10% Conservative: 83
Best Local Similarity: 21.12% Mismatches: 213

```

```

Query Match: 3.88% Indels: 141
DB: 2 Gaps: 19
US-09-787-016A-1 (1-2610) x I38073 (1-699)
QY 268 GACGACAAAGCGACCGCAGCAATGAGGAGGACCTAAGGCCATCAAAACCCACAGCAAA 327
DB 136 AspAspGluLysGlnLysGlnProValGlnLysGlyValLysProGlnAlaLys 155
QY 328 GAGTTTCAGGAAAAATGGGGTTTTCGAAGGACCACTATGCCCAAGCAGAGGGCGCAGGG 387
DB 156 AlaAlaLysAlaPro-----ProLysLysAlaLysSerSer 167
QY 388 GACGCGAGGCGTGAC---CCACTGGAGCGCCACCCACAGCAGCAGCTGGCGCTGTCC 444
DB 168 AspSerAspSerAspSerSerSerGluAspLupProLysAsnGln----- 183
QY 445 CTGCGGCGCAGTGGAGAGGACCCCAAGCGCAGCTGAGCGGTGGAGCAGTTCTGACCAATT 504
DB 184 -----LysProLysIleThr----- 188
QY 505 GCGGCGCGCGCGGAGGAGGAGCATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 564
DB 189 -----ProValThrValLysAlaGlnThrLysAlaPro 199
QY 565 TCCTGCCCCCGCACAGACCGCAGCAGCTCCGAGGCGCAGCGTGGAAAGCGCTTCTCTGAG 624
DB 200 ProLysProAlaAspAlaProLysIleAlaAsnGlyLysAlaAlaSerSerSerSer 219
QY 625 ACCAGAGCGCGCCCGCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 684
DB 220 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 239
QY 685 AAG-----GTCAAGAGGAGGAGTAC 705
DB 240 LysLysThrValProLysLysGlnValValAlaLysAlaProValLysAlaAlaThrThr 259
QY 706 CACGATGACACTCGATGATGACAGGATGCGCTGACCTTGAAGAGCTTCAGAAATCGC 765
DB 260 ProThrArgLysSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 279
QY 766 CTTCGACGAGGCGGAAACAGCAGGCGCC-----ACTGAGAGG 801
DB 280 LysProMetLysAsnLysProGlyProTyrSerTyrAlaProProProSerAlaProPro 299
QY 802 CCCCTGAAAGGATCCAGATGCGCTGCGGAGAAAGCCCGGAGGAGGCTCCCGCCGAG 861
DB 300 ProLysLysSerLeuGlyThrGlnProProLysLysAlaValGluLysGlnGlnProVal 319
QY 862 ACTGTGGGCTCCGAGCGCAGTACACTGTGGAGGCGCTCTGCCCAAGTAAGCAGAGCGCC 921
DB 320 GluSerSerGluAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 339
QY 922 GAGAACGATCAGGGGGTGTGTCCAGGCT-----GGGAAAGATGAC 963
DB 340 ProThr---LysAlaValSerLysAlaThrThrLysProProProAlaLysLysAla 358
QY 964 AGAGAGGTAGTTCGAGGGAAGCGGCTCAGGACATCAAGATGAGGAG----- 1014
DB 359 AlaGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 378
QY 1015 -----CCTGAGACTTGGGCGCAGCAGGAGCTGTAATGTGAGGGTTTACAGCCCAAGCC 1068
DB 379 LysProAlaGly-ThrThrLysAsnSerSerSerSerSerSerSerSerSerSerSer 398
QY 1069 CTGTATTGCTATTCGCGCAGCTCACAAACAGGTTTATGATTGCTGTGACGCTGT 1128
DB 398 oAlaValLysProAlaAla-AlaProLysGlnProValGly----- 411
QY 1129 GAAGATGTTTTCATGCGCATTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1188
DB 412 -----GlyGlyGln-LysLeuLeuThr 418

```


Db LysSerProLeuProSerLysGluAlaSerArgProThrSerVal----- 3264
QY 1087 CAGCCTCAACAACAGGTTTANGATTCTGTGACCGCTGTGAAGATGTTTCATGCG 1146
Db 3264 ----- 3264
QY 1147 GATTGTGCGCATTTCTCAGGCTCGAGGAGCCTTTTGAAGAAGAAATGGGAAGCATAT 1206
Db 3265 ---AlaLysSerValLysAspGluAlaGluLysSerLysGluLysSerArgAspSer 3283
QY 1207 ATCTGC-----CCAAATGCGACCATCTG 1230
Db 3284 ValAlaGluLysSerProLeuAlaSerLysGluAlaSerArgProAlaSerValAlaGlu 3303
QY 1231 CAATGTCAGGATCAGACT-----CAITCAGAAACGCGAGAT 1266
Db 3304 SerValGluLysGluAlaGluLysSerLysGluLysSerArgGluSerValAlaGlu 3323
QY 1267 CAGCAGGAAGCTAAATGGAGACCTGGAGATCTGTGATGCGACCATGTGTAAAGTATAGGA 1326
Db 3324 LysSerProLeuAlaTyrlLysGluAlaSerArgProAlaSerValAlaGluSerIleLys 3343
QY 1327 ACATATAGCAGAGCTAGCAGGAGCAAGCAAGGATTAAGGTTAGATTTCAGAAAGCT-- 1383
Db 3344 AspGluAlaGluLysSerLysGluLysSerArgGluSerValAlaGluLysSerPro 3363
QY 1384 ---GCAATCCCAAGTCCCAAGAGAACTCAAGATCTTCAGCCTGTGATAGAGCGCCT 1440
Db 3364 LeuAlaSerLysGluAlaSerArgProThrSerValAlaGluSerValLysAspGluAla 3383
QY 1441 GGTGCTCAAAATGTTTGGCCCGGCTGTCTACGTGCGGAGCCCGACCTCGGTGATC 1500
Db 3384 GluLysSerLys-----GluLysSerSerArgAspSerValAla 3396
QY 1501 TGCAGTAATGACTGTATCTCAACACGCGCGAGCAATGAAGTTCTTAAGCTCAGGT 1560
Db 3397 GluLysSerProLeuAlaSerLysGluAlaSerArgProAlaSerValAlaGluSerVal 3416
QY 1561 AAAGAACAGAGCAAGCTAAAGAAAGATGAAGTGAAG-----CCAGAGAGCC 1614
Db 3417 GluAlaGluLysSerLysGluLysSerArgGluSerValAlaGluLysSer 3436
QY 1615 AGTCTCCGAATGCGTCTCAGGAGGTATTAAATCTCTCTGTGACACAGAGACCA 1674
Db 3437 ProLeuAlaSerLysGluAlaSerArgProAlaSerValAlaGluSerValLysAspSer 3456
QY 1675 GTCTCCAGAAAAAAGAGACCAAGTCAAGAGCGAGTGTGTCTCCTCGCGAGTGA 1734
Db 3457 AlaGluLysSerLysGluLysSerArgGluSerValAla-----GluLysSerPro 3474
QY 1735 GCATCGGAGAGACAGCTTGTGAGAGCAGCAGCGCTGCTGCGGAGCGATCAC--- 1791
Db 3475 LeuAlaSerLysGluAlaSerArgProAlaSerValAlaGluSerValLysAspGluAla 3494
QY 1792 -----AATTACAATGAGTAAAGCAGCAAGAGAGTGTCTCCTCGCGC 1836
Db 3495 GluLysSerLysGluLysSerArgGluSerValAlaGluLysSerPro 3511
RESULT 6
JC5662
N;Alternate names: HRP-2
C;Species: Mus musculus (house mouse)
C;Date: 17-Nov-1997 #sequence_revision 17-Nov-1997 #text_change 05-Nov-1999
C;Accession: JC5662
R;Izumoto, Y.; Kuroda, T.; Harada, H.; Kishimoto, T.; Nakamura, H.
Biochem. Biophys. Res. Commun. 238, 26-32, 1997
A;Title: Hepatoma-derived growth factor belongs to a gene family in mice showing signifi
A;Reference number: JC5660; MUID:97445118; PMID:9299445
A;Accession: JC5662
A;Molecule type: mRNA
A;Residues: 1-669 <I>ZU>
A;Cross-references: DDBJ:D63850; NID:G2558500; PIDN:BAA22896.1; PID:di023766; PID:G25585

C;Comment: This protein translocates to the nucleus and directly functions in mitogenic
F:1-93/Region: hath #status predicted
F:321-363/Region: nuclear location signal

Alignment Scores: 0.000608 Length: 669
Pred. No.: 180.50 Matches: 131
Score: 33.44% Conservative: 79
Percent Similarity: 33.44% Mismatches: 193
Best Local Similarity: 20.86% Indels: 225
Query Match: 3.85% Gaps: 29
DB:

US-09-787-016A-1 (1-2610) x JC5662 (1-669)

QY 220 AGGATTTTCAGGAAAAAGTGTCTCAAGCTTTTCAGTGTGTGAGCAGCATGTGACACGAAAGGC 279
Db 171 ArgArgAlaSerSerAspLeuAlaSerValSerProSerGluLysSerGlu 190
QY 280 GACCGGCAATAGAGGAGGACCTTAAGCCCATCAAAACCCACCACCAAGAGTTTCAGGAAA 339
Db 191 SerProSerGluSerGlu-----LysThrSerAspGluAspPhe----- 203
QY 340 ACATGGGGTTTTGGAAGGACCACTATCGCCAAAGCGAGGCGCGAGGGGACGCGGAGGCT 399
Db 204 -----ThrProGluLysLysThrAlaAla----- 211
QY 400 GACCCACTGGAGCGCCACCCACAGCAGCAGTGTGGCCCTGTCTCCTCGCGCGCAGTGGG 459
Db 212 -----ArgProProArgArgGlyProLeuGly-----Gly 221
QY 460 AGGAGCCCAAGCGCAGCTGAGCGCGTGTGAGCAGTTCCTGACCATTCGCGCGCGCGCGC 519
Db 222 ArgLysLysLysLysValProSerAlaSerAspSerAspSerLysAlaAspSerAspGly 241
QY 520 AGGAGGACATGCTCTCTCCTCGAGGATTCGTGTGAGCCACGCTCTCTGCGCGCCACA 579
Db 242 AlalysGluLysProValVal-----ThrAlaGluProSer-----ProSerSer 256
QY 580 GACCCGAGACAGCTCTCGAGGCGAGCTGTGAAAGCGCTTCTGACACCAAGAGCGCGCC 639
Db 257 SerSerSerSerSerSerSerSerSerSerSerSerSerSerValLysLysProPro 276
QY 640 CAGTCTGCTTCCACAGCTGTGAAG----- 663
Db 277 ArgGlyArgLysProAlaGluLysProProLysProArgGlyArgArgProLysPro 296
QY 664 GAACGACAGCTCTTCTGAAAGGTGAAGAGGGGATGACCAAGATGACACCTCCGAT 723
Db 297 GluArgProProSerThr-----SerSerSerSerAsp 306
QY 724 AGTGACAGCGATGGCTGACCTTGAAGAGCTTCAGAAT----- 762
Db 307 SerAspSerAspSerGlyLysValAspArgLysSerGluTrpLysArgArgAspGluGlu 326
QY 763 -----CGCTTCGAGGAGCGGACAGGAGCCCTCAGAGGCGCCCTG 807
Db 327 ArgArgGluLysGluAlaArgArgArgGluGluGluGluGluGluGluGluGlu 346
QY 808 AAAGGATCCACAGTCTGCTGCGGAAGAGAGCGCGGAGGAGGCGCGCGCGAGACTGTG 867
Db 347 ArgGlu-----GluLysGluLysGluLysGluLysGluLysGluLysGluLysGlu 364
QY 868 GGCTCCGAGGCGCAGTGACACTGTGTGAGGCGCTCTGCGCCAGTAAGCAGGAGCCCGAGAAC 927
Db 365 GlySer-----SerGlyGluLysGluLysGluLys 373
QY 928 GATCAGGGGGTTTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGTAAGTTTGAGGGAAG 987
Db 374 GluLysProVal-----LysLysArgSerArgLysAlaArgGlyArg 387
QY 988 GCGGCTCAGGACATCAAGATGAGAGCT---GGAGACTTGGGCGCA----- 1032
Db 388 GlyThrProSerSerSerAspSerGluProGluGlyGluLysGluLysGluLysLys 407

1033 QY -----CCGAAGCCTGAATGTGAGGGTTCAGACCCCAAGCCCTGTAATTCATTTCCCGC 1086
 408 Db LeuAlaLysSerGlnLeuProGlySerGluSer-----AlaArg 421
 1087 QY CAGCCTCAACAACAGGTTTATGATTGCTGTGACCGCTGTGAAGAATGTTTCATGCC 1146
 422 Db LysPro----- 423
 1147 QY GATTGTGTGGCATTCTTCAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGAAGACTAT 1206
 424 Db -----GlyGlnLysGluLysArg----- 431
 1207 QY ATCTGCCCAACGCAACCATCTTCTGCAAGTCAGGATGAGACTCATTCAGAAACGGCAGAT 1266
 432 Db -----Pro 432
 1267 QY CAGCAGGAAGCTAAATGGAGACCTGGAGATGCTGATGGCACCAGATTGTACAGTATAGGA 1326
 433 Db AspGluLysProAlaArgProValLysValGluArgThr----- 446
 1327 QY ACAATAGAGCAGGAAGTCTACGCAAGACCAAGGGATAAAGGTAGAAATTGAGAAGCTGCA 1386
 447 Db -----ArgLysArgSerGluGlyLeuSerLeuGluArgLysGlyGluLysLys 463
 1387 QY AATCCAGTGGCAAGAAACTCAAG-----ATCTCCAGCCTGTG 1428
 464 Db GluProSerValGluArgLeuGlnLysLeuHisSerGluLeuLysPheAlaLeuLys 483
 1429 QY ATAGAGCGGCTGTGCTCAAAATGATTATGGCCCC-----GGGTCTCTGCAC 1476
 484 Db ValAspAsnProAspValArgLysCysLeuSerAlaLeuGluLeuGlyThrLeuGln 503
 1477 QY GTGGCG-----CAGCCCGACTCGTG----- 1497
 504 Db ValThrSerGlnLeuGlnLysAsnThrAspValValAlaThrLeuLysLysIleArg 523
 1498 QY ---TACTGCAGTAACTGCTATCTCTCAACACGCGCA----- 1533
 524 Db ArgTyrLysAlaAsnLysAspValMetAlaLysAlaAlaGluValThrArgLeuLys 543
 1534 QY -----GCGACAAATGAAGTTTCTAAGCTCAGTAAAGAA 1566
 544 Db SerArgValLeuGlyProLysValGluAlaLeuGlnLysValAsnLysAlaGlyAlaGlu 563
 1567 QY CAGAGCCAAAGCTTAAGAAAGATCAAGATGACGACAGAGCCAGCTCTCCGAA 1626
 564 Db LysGluArgAlaAspAsnGluLysLeuGluGlnPro----- 576
 1627 QY TGGGGTCTCAGCAGGTATTAAATCTTCTCTGTGCAC-----AAGAGACCAAGCT 1677
 577 Db ---GlyGluGlnAlaProArgGluLeuAlaGluAspGluProSerThrAspArgSerAla 595
 1678 QY CCAGAAAAAAGAGACCACAGTAGAGAGCGAGTGTGTCTCCGCGGAGTGAAGCA 1737
 596 Db ProValAsnGlyGluAlaThrSerGlnLysGlyLysAsnMetGluAspArgAlaGlnGlu 615
 1738 QY CTGGGAAGCAAGCAGCTTGTGAGAGCAGCACCGCGTGTGGCGAGCGATCAC----- 1791
 616 Db AspGlyGlnAspSer-----GluAspGlyProArgGlySerSerGluLeuHis 633
 1792 QY -----AATTCAATGCAAGTAAAGCA-----GAA 1815
 634 Db AspSerProArgAspAsnSerAspProAlaLysProGlyAsnGluArgGlnAspHisGlu 653
 1816 QY AAGACTGCTCTCCCTCCGCTCA 1839
 654 Db ArgThrArgLeuAlaSerGluSer 661

RESULT 7

A43427

neurofilament triplet H1 protein - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 10-Dec-1999

C:Accession: A43427

R:Soppet, D.R.; Beasley, L.L.; Willard, M.B.

J. Biol. Chem. 267, 17354-17361, 1992

A:Title: Evidence for unequal crossing over in the evolution of the neurofilament polype

A:Reference number: A43427; PMID:92381055; PMID:1512270

A:Accession: A43427

A:Molecule type: DNA

A:Residues: 1-606 <SOP>

A:Cross-references: GB:M94315; MID:g164990; PIDN:AAA57152.1; PID:g601930

A:Experimental source: brain

A:Note: sequence extracted from NCBI backbone (NCBIN:112010, NCBI:P:112011)

C:Superfamily: neurofilament triplet H protein

C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Alignment Scores:

Pred. No.:	0.000647	Length:	606
Score:	180.00	Matches:	106
Percent Similarity:	34.9%	Conservative:	77
Best Local Similarity:	20.2%	Mismatches:	211
Query Match:	3.84%	Indels:	129
DB:	2	Gaps:	19

US-09-787-016A-1 (1-2610) x A43427 (1-606)

QY	312	CAAAACCCACCAAGAGTTTCAGAAACATGCGGTTTTCGAAGACCACTATCGCCAA	371
----	-----	---	-----

Db	10	GlnThrGluGlnValThrGluVal-----ThrGluGluGluLysGlu	27
----	----	---	----

QY	372	CGCAGAGGCGCAGGAGCGCGAGGCTGACCCATCGAGCGCCACCCACCA	425
----	-----	--	-----

Db	28	Ala-----LysGluGluGlyGlyGluGluGluAlaLysSerProThrGluGly	45
----	----	---	----

QY	426	GCAGCAGCTGGGCTCTCCCTGGCGGCGAGTGGAGGCGCCAGCGCCTGAGCGCT	485
----	-----	---	-----

Db	46	GlyAlaAlaSerProGluGluAlaLys-----	55
----	----	----------------------------------	----

QY	486	CGAGCAGTTCTTGACCATTCGCGCGCGCGAGGAGGAGCATCTCTCTCCCTGA	545
----	-----	--	-----

Db	56	-----SerProAlaGluAlaLys-SerProValLys---G	66
----	----	--	----

QY	546	GAATCTGTGTAGCCACGCTCTGCCCCCAGACAGCGCGAGACAGCTCCGAGGCGAG	605
----	-----	---	-----

Db	66	uGluAlaLysSerPro-----AlaGluAlaLysSerProAlaG	79
----	----	---	----

QY	606	CCTGAAAGCGCTTCTGAGACCAAGCGCCCGAGTCTGCTTCCACAGCTGTGAAGA	665
----	-----	--	-----

Db	79	uAlaLysSerProAlaGluAlaLysSer---ProGluLysAlaLysSerProValLysG	98
----	----	---	----

QY	666	ACGACAGCTTCTCTGAAAAGGTGAAGAGGGGATGACACGATGACACCTCCCATAG	725
----	-----	---	-----

Db	98	uGluAlaLysSerProGluLysAlaLysSerProValLysGluGluAlaLysSerProAl	118
----	----	--	-----

QY	726	TGACAGCGATGCTGACCTTGAAGAGCTTCAGAATCGCTTCGACGAGCGGGAACA	785
----	-----	--	-----

Db	118	aGluAla-----LysSerProGluLysAlaLysSerProAlaGluAl	132
----	-----	---	-----

QY	786	GGAGCCCTAGAGGCGCCCTGAAAGGATCCAGAGTCCGCTGCGGGAAGAGCGCGGA	845
----	-----	---	-----

Db	132	alysSerProGluLysAlaLysSerProValLysGluGluAlaLysSerProGluLysAl	152
----	-----	--	-----

QY	846	GGAGGTCCTCCGCGACACTGTGGCTCCGAG---GCCAGTACACTGTGGCGGCTCT	902
----	-----	---	-----

Db	152	alysSerProAlaGluAlaLysSerProGluLysAlaLysSerProAlaGluAlaLysSe	172
----	-----	--	-----

QY	903	GCCTAGTAAG-----CAGGAGCCCGAGAACGATCAGGG	935
----	-----	--	-----

Db	172	rProGluLysAlaLysSerProValLysGluGluAlaLysSerProGluLysAlaLysSe	192
----	-----	--	-----

QY	936	GGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGTANGTTGGAGGGAAGCGGCTCA	995
----	-----	--	-----

Db	192	rProValLysGluGluAlaLysSerProAlaGluAlaLysSerProGluLysAlaLysSe	212
----	-----	--	-----

QY 996 GGACATCAAGATGAGGAGCTGGAGACTTGGGCGGACGACCTGATGTTGAGGTTA 1055
Db : : : : :
QY 212 rProValylsGluLysSerProGluLysAlaLysSerPro----- 227
Db : : : : :
QY 1056 CGACCCCAACCCCTGATTCATTTGCGCCAGCTCAACAACAGGTTTATGATTTG 1115
Db 227 ----- 227
QY 1116 CTGTGACCGCTGTGAAGATGTTTCATGCGGATGTTGTGGCATTTCTGAGGCTCGAGG 1175
Db : : : : :
QY 228 -----AlaGluAlaLysSe 232
Db : : : : :
QY 1176 GAGGCTTTTGAAGGAATGGGAAGACTATATCTGCCCAACTGCACCATTTCTGCAAGT 1235
Db : : : : :
QY 232 rProValylsGlu-----GluAl 238
Db : : : : :
QY 1236 GCAGGATGACACTCATTCAGAACGGCAGATCAGCAGGAGTAAATGAGACTCGAGA 1295
Db : : : : :
QY 238 aLysSerProGluLysAlaLysSerProGluLysGluLysAlaLysSerProGluLysAla 257
Db : : : : :
QY 1296 TGCTGTGATGGCAGCCGATTTGTAAGATATAGGACAAATAGACAGAGTCTAGCGAAGACCA 1355
Db : : : : :
QY 257 uAlaLysSer-----ProGluLysAlaLysSerProGlu 268
Db : : : : :
QY 1356 AGGATTAAGGATGATTCAGAAAGCTGCAATCCAGTGGCAAGAACTCAAGAT 1415
Db : : : : :
QY 268 uLysAlaLysSerProValGluValLysSerProAlaGluAlaLysSerProGluLysAla 288
Db : : : : :
QY 1416 CTTCCAGCTGTG---ATAGAGGCGCTGTGCTCAAAATGATTTGGCCCGGCTG 1472
Db : : : : :
QY 288 aLysSerProValylsGluLysSerProGluLysAlaLysSerProValylsGlu 308
Db : : : : :
QY 1473 TCAGTGGCGGCGGAGCTGGTGTACTGCAATGATGATCTCTCAACACGCGC 1532
Db : : : : :
QY 308 uGluAlaLysSerProGluLys-----AlaLysSerProValylsGluLysAla 325
Db : : : : :
QY 1533 AGCGACAATCAAGTTCTTAAGCTCAGGTAAAGAACAGACCAAGCCT---AAAGAAA 1589
Db : : : : :
QY 325 sSerProGluLysAlaLysSerProValylsGluLysAlaLysSerProGluLysAla 345
Db : : : : :
QY 1590 GATGAGATGAGCCAGAGAGCCAGCTCTCCGAATGCGGTGCTCAGCAGGATTA 1649
Db : : : : :
QY 345 sSerProValylsGluLysAlaLysSerProGluLysAlaLysSerProValylsGlu 365
Db : : : : :
QY 1650 AATCTCTTCTGTGCACAGAGACCCAGCTCCAGAAAAAAGAGACACAGTGAAGAGGC 1709
Db : : : : :
QY 365 uAlaLysSerProGluLysAlaLysSerProGluLysAlaLysSerProValylsGlu 385
Db : : : : :
QY 1710 AGTGGTGTCTCCGCGGAGTGAAGCAGCTCGGGAAGGAGCAGCTTGTGAGCAGCAC 1769
Db : : : : :
QY 385 uAlaLysSerProGluLysAlaLysSerProValylsGlu-----GluAlaLysSe 402
Db : : : : :
QY 1770 GCGTGTGCGGAGCGATCACAATTACATGAGTAAAGCCAGAAAGACTCTGCTCC 1829
Db : : : : :
QY 402 rProGlu-----LysAlaLysSerProValylsGluLysAla 415
Db : : : : :
QY 1830 CTGCGCG 1836
Db : : : : :
QY 415 sSerPro 417
Db : : : : :
RESULT 8
151116
NF-180 - sea lamprey
C:Species: Petromyzon marinus (sea lamprey)
C>Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: I51116
R:Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A>Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A:Reference number: I51116; MUID:95287814; PMID:7770000
A:Accession: I51116
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA

A:Residues: 1-1110 <JAC>
A:Cross-references: EMBL:U19361; NID:G632548; PIDN:AAA80106.1; PID:G632549
C:Superfamily: neurofilament triplet H protein
Alignment Scores:
Pred. No.: 0.00118 Length: 1110
Score: 176.50 Matches: 113
Percent Similarity: 35.60% Conservative: 86
Best Local Similarity: 20.21% Mismatches: 235
Query Match: 3.76% Indels: 125
DB: 2 Gaps: 23
US-09-787-016A-1 (1-2610) x I51116 (1-1110)
QY 253 GTTGGAGCAGGTATGAGCAGCAAAAGCGCCGAGCAATGAGGAGCACTTAAGCCATC 312
Db ValSerAlaGluLysAspGluGluGluGluGluGluGluGluGluGluGluGluGlu 500
QY 313 AAACCCACCAGCAAGATTCAGGAAACATGGGGTTCGAAGGACCCTATCGCCAAG 372
Db : : : : :
QY 501 AlaGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 513
Db : : : : :
QY 373 CGAGAGGGCGCAGGAGCGCGGAGCTGACCCACTGGAGCCGCCACCCGCCAGCAGCAG 432
Db LysGluGlyGluAlaGluAlaGluGluGluGluGluGluGluGluGluGluGlu 533
QY 433 CTGGGCGCTTCTCCTCGCGCGCAGTGGGAGCGACCAAGCGCAGCTGAGCGCTGGAGCAG 492
Db : : : : :
QY 534 GluGluAlaGluLysGluGluAlaGluGluGluGluGluGluGluGluGluGlu 553
QY 493 TTCCTGACCATTCGCGCGCGCGCGCAGGAGGATGCTCTCTCTCCCTGGAGGATTC 552
Db : : : : :
QY 554 -----GluGluGluAlaGluGluGluGluGluGlu 562
QY 553 GGTGAGCCCATCTCTGCGCGCGCGCAGCAGCGCGGAGCAGCTCCGAGGCGCAGCTGGA 612
Db AlaGluAlaGluGlyGluGluAlaGluGluGluGluGluGluGluGluGluGlu 582
QY 613 AGCGCTTCTGAGCAGAGCGCGCGCGCGCTCTGCTCCACAGCTGTGGAAGCAACGACCA 672
Db LysAlaGluAlaAla-----GluAlaLysAlaGluGluGluGluGluGlu 597
QY 673 GCTCTTCTGAAAAGGTGAAGAGGGATGACACGATGACACCTCCGATGATGACAC 732
Db AlaGluAlaGluGlu-----GluGluGluGluGluGluGluGluGluGlu 613
QY 733 GATGCGCTGACCTTGAAGAGCTTCAGATCGCCTT----- 768
Db : : : : :
QY 614 GluAlaGluThrLysGluGluValGluGluAlaGluAlaGluValGluGluGluGlu 633
QY 769 -----CGCAGGAAGCGGAAACAGAGGCCCTGAGAGGCCCTGAAAGGGATCCAGAT 822
Db AlaGluGluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGlu 653
QY 823 CCGCTCGGAGAAAGCGCGGAGGAGGTCCCGGAGACTGTGGCTCCGAGGCCACT 882
Db LysThrGluGluAlaGluValGluGluGluGluGluGluGluGluGluGluGlu 673
QY 883 GACACTGTGAGGCGGTCTCTGCCAGTAAGCAGGAGCGCGGAGACGATCAGGGGTGTG 942
Db : : : : :
QY 674 GluGluAlaGlu-----GluGluAlaGlyGluGluGluGluGlu 686
QY 943 TCCGAGCTGGGAAATGATCAGACAGAGATAGTTGAGGGAAGAGCGCTCAGGAC--- 999
Db : : : : :
QY 687 AlaGluSerLysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 706
QY 1000 -----ATCAAGATGAGGAGCTTCGGAGACTTCGGCGCGAGCCGAGGAGCTGAA 1044
Db : : : : :
QY 707 AlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 726
QY 1045 TGTGAGGTTACAGCCCAACCGCTCTGATTGCTATTCGCTTGGCGGAGCTCAGCAACAGG 1104
Db : : : : :
QY 727 AlaGluAlaGluGluGluAlaAla----- 734

QY 1105 TTTATGTTGCTGTGACCGCTGTGAGGATGTTTTCATGCGGATGTTGCGCAATTTCT 1164
Db -----LysSerGluGlu-----GluAlaAlaGlu 742
QY 1165 GAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGAAGACTATATCTGCCAAACTGCACC 1224
Db -----LysSerGluGlu-----GluAlaAlaGlu 755
QY 1225 ATTCTGCAAGTGCAGGATGAGCTCATTCAGAAACGCGCATCAGCAGGAAGCTAATGG 1284
Db -----GluAlaValGluGluThrGluAlaAlaThr---GluGluAlaGluAlaLysGlu 772
QY 1285 AGACCTGGAGATGCTGATGCGACCGATTTGACAGTATAGAACATAGACAGAGAGTCT 1344
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1345 AGCGAA-----GACCAGGAGTAAGAGGTAGAAATTCAGAAAGCTGCAATCCA 1392
Db -----LysSerGluGlu-----GluAlaAlaGlu 742
QY 1393 AGTGGCAAGAGAACTCAAGATCTCCAGCCTGTGATAGAG-----GGCCTGGT 1443
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1444 GCTCAAAATGTTATGGCCCGGCTGTCTGCTGCGCGCAGCCGACTCGGTGTTACTGC 1503
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1504 AGTAATGACTGTATCTCAACACCGCGCAGCAATGAGTTTCTAAGCTCAGGTAA 1563
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1564 GAACAGAGCAAG-----CCTAAGAAAGATGAAGTGAAGCAGAGAGCCAGT 1617
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1618 CTTCGGAATGCGTGTCTGAGGAGGATTAATAATCTTCTGTGCAAGAGACCACT 1677
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1678 CCAGAAAGAGAGACCAAGTGAAGAGGAGGAGTGTGCTGCGCGAGTGA--- 1734
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1735 GCATCGGAGAGAGAGCTGTGAGAGCAGCAGCCGCTGCTGCGGAGCGGATCAAT 1794
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1795 TACAATGCAATGAGCAGAGCAAGAGACT-----CCTGCTCCTCGCGCTCA 1839
Db -----LysSerGluGlu-----GluValLysGlu 785
QY 1839 TACAATGCAATGAGCAGAGCAAGAGACT-----CCTGCTCCTCGCGCTCA 1839
Db -----LysSerGluGlu-----GluValLysGlu 785

A;Residues: 32-180,'Q',181-734 <ME3>
A;Cross-references: GB:M94287; PIDN:AAA41718.1; PID:g205750
R;Meier, U.; Blobel, G.
Submitted to the EMBL Data Library, May 1992
A;Description: Noppl40 shuttles on tracks between nucleolus and cytoplasm.
A;Reference number: S27889
A;Accession: S27890
A;Molecule type: mRNA
A;Residues: 32-734 <ME2>
A;Cross-references: EMBL:M94288; NID:g205751; PIDN:AAA41719.1; PID:g205752
A;Accession: S27889
A;Molecule type: mRNA
A;Residues: 32-180,'Q',181-734 <ME12>
A;Cross-references: EMBL:M94287; NID:g205749; PIDN:AAA41718.1; PID:g205750
C;Superfamily: nucleolus-cytoplasm shuttle phosphoprotein
C;Keywords: phosphoprotein

Alignment Scores:
Pred. No.: 0.00163 Length: 734
Score: 174.00 Matches: 132
Percent Similarity: 35.42% Conservative: 77
Best Local Similarity: 22.37% Mismatches: 250
Query Match: 3.71% Indels: 131
DB: 25 Gaps: 25

US-09-787-016A-1 (1-2610) x B42680 (1-734)

QY 172 TCTGCTTGACCAAGAGGTTTCGTCCTCCAGGCTTTTGGTGTATTTAGGATTCAGGG 231
Db 86 SerPheTrpLeuLysSerThrLysAlaPro-----LysValLysLeuGlnSerAsnGly 103
QY 232 AAAAGTGTCAAGCTTTCAGTGTGGAGCAGGTATGAGCAGCAAGGCGACCCGCAAT 291
Db 104 ProValAlaLysLysLysLysGluThrSerSerSerSerSerSerSerSerSerSer 123
QY 292 GAGGAGGACCTTAAGCCATCAACCCAGCAGCAAGAGTTCAGGAAACATCGGGGTTT 351
Db 124 GluGluGluAspLysAlaGlnValProThrGlnLys----- 135
QY 352 CGAAGACCACTATGCCAAGCA-----GAGGCGCAGGCGACCGGAGGCT 399
Db 136 ---AlaAlaAlaProAlaLysArgAlaSerLeuProGlnHisAlaGlyLysAlaAla 154
QY 400 GACCCACTGCGCGCCGCCACCCAGCAGCAGCTGGGCTCTCCCTGCGGCGCAGTGGG 459
Db 155 LysAlaSerGluSerSerSerSerGluGlu-----SerSerGluGluGlu 170
QY 460 AGGCGACCCAGCAGCTGAGCGCTGGAGCAGTCTCTGACCATTCGCGGCGCGCGGC 519
Db 171 GluLysAspLysLysLysProValGlnLysAlaValLysProGlnAlaLysAlaVal 190
QY 520 AGGAGGAGCTCTCTCTCTCTGAGGATTCGTGAGCCCGCAGCTCTCTCTCTCTCT 579
Db 191 Arg-----ProProLysLysLysLysLysLysLysLysLysLysLysLysLys 208
QY 580 GACGCGGAGCAGCTCTCCGAGGCGCGTGGAAAGCGCTTCTGAGACCAAGAGCGCC 639
Db 209 GluAspGluAlaProGlnThrGlnLysProLysAlaAlaAlaAlaAlaAlaAlaPro 228
QY 640 CAGTCTGCTTCCAGCAGCTGTGAGGAGCAGCAGCTCTCTCTCTCTCTCTCTCTCT 690
Db 229 ThrLysAlaGlnThrLysAlaProAlaLysProGlyProProAlaLysAlaGlnProLys 248
QY 691 -----AAAGAGGAGGATGACACGATGACACCTCCGATAGTGACGAGGATGGC 738
Db 249 AlaAlaAsnGlyLysAlaGlySerSerSerSerSerSerSerSerSerSerSerSer 268
QY 739 CTGACCTTGAAGAGCTTCAGAACTCGCTTCGAGGAACCGGAGACAGAGCCCACTGAG 798
Db 269 -----SerGluGluGluLysLysAlaAla 276
QY 799 AGGCGCTGAGAGGATCCAGAGTCCGCTCGGAGAGCGCGCGGAGGAGGCTCCCGCC 858
Db ----- 858

```
Db 277 AlaProLeuLys -----LysThrAlaProLys 285
Qy 859 GAGACTGTGGCTCCGAGGCGCAGTACACTGTGGAGGCGTCTGCCAGTAAGCAGGAG 918
Db 286 LysGlnValValAlaLysAlaProValLysValThrAlaLalProThrGlnLysSerSer 305
Qy 919 CCGAGAACCATCAGGCGGTGTGTCCAGGCTGGGAAGATGACAGAGAGATAGTTG 978
Db 306 SerSerGluAsp -----SerSerGluGluGluGluGlnLysLysPro 321
Qy 979 GAGGGAAGCGGCTCAGGACATCAAGATGAGGAGCTGGAGACTTGGGCCCGCAGGAG 1038
Db 322 MetLysLysAlaGlyProTyrSerSerValProProSerValSerLeuSerLys 341
Qy 1039 CTGAATGTGAGGTTACGACCCCAAGCCCTGTATTGCTATTGCCCGCAGCCCTCAAC 1098
Db 342 LysSerValGlyAlaGlnSerProLysLysAla ---AlaAlaGlnThrGlnProAlaAsp 360
Qy 1099 AACAGGTTATGATTCTGTGACCGCTGTGAAGATGGTTTCATGCGGATTTGTGGGC 1158
Db 361 Ser -----SerAlaAspSerSerGluGlu ----- 368
Qy 1159 ATTCTTGAGGCTCAGGAGGCTTTTGGAAAGAAATGGGAAGACTATATCTGCCCAAAC 1218
Db 369 ---SerAspSerSerSerGlu -----GluGluLysLysThrProAla 381
Qy 1219 TGCACCATCTGCAAGTGCAGGATGAGACTCATTCAGAACGGCAGATCAGCAGAGCT 1278
Db 382 LysThrValVal -----SerLysThrProAlaLysProAlaProValLysLysLys 398
Qy 1279 AATGGAGACTGCGAGATGCTGATGCGCACCGATGTACAAAGTATAGGAACATAGAGCAG 1338
Db 399 AlaGluSerSerSerAspSerSerAsp ----- 409
Qy 1339 AAGTCTAGCAGAACCAAGGATTAAGGTAGAAAGTGAAGTCAAAATCCAGTGGC 1398
Db 410 ---SerSerGluAsp -----GluAlaProAlaLysProValSer 421
Qy 1399 AAGAAGAACTCAAGATCTCCAGCT ---GTGATAGAGCGCTGTGCTCAAAATGT 1455
Db 422 AlaThrLysSerProLysLysLysProAlaValThrProLysProProAlaAlaLysAla 441
Qy 1456 ATGGCCCCGGTCTGTACGTGGCGCAGCCGACTCGGTGTACTGCACTAATGACTGT 1515
Db 442 ---ValAlaThrProLysGlnProAlaGlySerGlyGln--L 454
Qy 1516 ATCTCTAAACACGCGCAGCAGCAAT ---GAAGTTCTAAGCTCAGGTAAA 1563
Db 454 ysProGlnSerArgLysAlaAspSerSerSerSerGluGluLysSerSerSerGlu 473
Qy 1564 GAACAGAACCCAAAG ---CCTAAAGAAAGATGAAGATCAAGCCAGAG 1608
Db 474 GluGluAlaThrLysLysSerValThrProLysAlaArgValThrAlaLys---Ala 492
Qy 1609 AAGCCAGTCTTCG ---AATGCGGTGCTCAGCAGGTATTAAATCTCTTGTGTCAC 1665
Db 493 AlaProSerLeuProAlaLysGlnAlaProArgAlaGlyLysSerSerSerAspSer 512
Qy 1666 AAGACAGCAGCTCCAGAAAAAAGAGACACACAGTGAAGAGGCGGTGGTCCCTGCG 1725
Db 513 GluSerSerSerGluGluGluLysLysThrProProLysPro-----ProAla 529
Qy 1726 CGAGTGAAGACTCGG ---AAGGAAGCAGCTTGT 1758
Db 530 LysLysLysAlaAlaGlyAlaAlaValProLysProThrProValLysLysAlaAla 549
Qy 1759 GAGACAGCAGCGCTGCGGCGCAGCGATCACAAT-----TACAATGCAAGTAAAGCCA 1812
Db 550 GluSerSerSerSerSerSerSerSerSerGluAspSerSerGluGluLysLysPro 569
Qy 1813 GAAAGACTGTGCTCTCCGCG 1836
Db 570 LysSerLysAlaThrProLysPro 577
```

RESULT 10

```
T21430
Hypothetical protein F26H11.3b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21430
R:Barlow, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19421
A:Accession: T21430
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-510 <WIL>
A:Cross-references: EMBL:Z81515; PIDN: CAB04195.1; GSPDB: GN000020; CESP: F26H11.3b
A:Experimental source: clone F26H11
C:Genetics:
A:Gene: CESP:F26H11.3b
A:Map position: 2
A:Intons: 116/1, 147/3; 288/2; 348/2; 392/1
C:Superfamily: Bromodomain homology
F:371-426/Domain: Bromodomain homology <BRO>
```

```
Alignment Scores:
Pred. No.: 0.00263 Length: 510
Score: 170.50 Matches: 92
Percent Similarity: 31.29% Conservative: 46
Best Local Similarity: 20.86% Mismatches: 145
Query Match: 3.64% Indels: 158
DB: 2 Gaps: 19
```

US-09-787-016a-1 (1-2610) x T21430 (1-510)

```
Qy 60 CTGGGGAGCTTACTCCACGGGACAGCGCTCTAGATAATCTGAGTTGTTGAAATAACGAA 119
Db 16 ProSerGlnTyrCysPro---SerThrValLeuSerAsnLeuThrLeu----- 30
Qy 120 GCCTGT---TACTCGTGAACAGCTGACACAGTGTGTGTGTGAGCCTGCTGTCTGC 176
Db 31 SerCysProTyrGlnAspGlnGluAlaAspAlaSerLeuGluIleSer----- 46
Qy 177 TTGGACCCAGAGGTTCTGCTCCAGGGTTTGGTGTGATTATTAGATTCTAGGGAAAAG 236
Db 47 -----AspPheAlaGluVal 51
Qy 237 TGTCCAAGCTTT-----CAGTGTGGAGCAGGTATGAGACACAAAGCGCACCC 284
Db 52 LeuProLysPheGluSerAlaGluGlnAspTyrAsnSerPheGlyTyrLeuLeu----- 69
Qy 285 GAGCAATGAGGAGGACCTTAAGGCCATCAACCCACCAAGAGATTGAGGAAACATG 344
Db 70 -----AsnGluGln-ProGlyThrSerSerAspLysThrThr 82
Qy 345 GGTTTTCGAAGACCACTATCGCCAGGAGAGGCGGCGGAGGCGGAGCGGCTGACCC 404
Db 82 rProLysLysIleThrValPheGlnLys-----Pr 93
Qy 405 ACTGGAGCGCCACCCACCCACAGCAGCTGGGCTGTCTCCCTCGCGCGAGTGGAGGCA 464
Db 93 oValGluPro-----IleGlyLysG 100
Qy 465 GCCCAAGCGCACTGAGCGGCTGGGAGTCTCTGACCATTTGCGCGCGCGCGC-----GG 518
Db 100 yProArgArg-----ArgArgArgCysAlaAs 109
Qy 519 CAGGAGGAGCATCTGTCTCTGAGGAGTCTGCTGAGCCACCATCTCTGCCGCCGCCAC 578
Db 109 pArgGluIleSerGluLeuAlaAlaLysProLysAlaGluValLysLysGluValLeAs 129
Qy 579 AGACGCGGACACAGCTCCGAGGCGGCTGGAAAGCGCTTCTGAGACACAGAGCGGCC 638
Db 129 nProAlaAspIleThrLeuGlyAspThrTyrAspTyrValLysGluGlnLysProTh 149
```


QY 1201 GACTATATCTGCCCAACTGCACCATTTCTG-----CAAGTGCAGGATGAGACTCAT 1251
 Db 642 -----LeuSerProAlaValLysThrLeuProLysLysAlaGluSerSerSer 659
 QY 1252 TCAGAAACGGCAGATCAGCAGCAAGCTAAATGAGAGCTCGAGATGCTGATGGCAGCAT 1311
 Db 660 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 679
 QY 1312 TGTCACAGTATAGGAACA-----LysSerLysGln 713
 Db 680 AlaThrProValAsnThrLysAlaProAlaGlnAsnLysAlaSerLysAlaSerCysSer 699
 QY 1330 ATAGAGCAGAAGCTCTAGCAGAGACCAAGGATAAAGGGTAGATTTGAGAAAGCTGCAAT 1389
 Db 700 AspSerAspSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 713
 QY 1390 CCAAGTGGCAGCAAGAAACTCAAGATCTCCAGCTGTGTAGAGCGCCTGGTGCCTCA 1449
 Db 714 ProThrGlyLys-----SerProAlaAlaLysAlaThrAlaProPro 727
 QY 1450 AAATGATATGGCCCGGTGTGTCACTGCGCAGCGCCGACTCGGTGTACTGCAAT 1509
 Db 728 Lys-----LysAsnProValAlaValAsnLysAspLysProSerSerSerSerSer 746
 QY 1510 GACTGTATCTCAACACCGCCGACGACATGAATTTCTAAGCTCAGGTAAAGACAG 1569
 Db 747 -----SerSerGlyAspSerGlu 752
 QY 1570 AAGCCAAAGCCTAAGAA-----AAGATCAAGATGAAGCCAGAG 1608
 Db 753 LysGlnLysProLysGlnAlaAlaAlaLysAspValLysGlnGlyAlaLysAlaAla 772
 QY 1609 AAGCCAGTCTCCGAATGGGTGCTCAGCAGGATTAATAATCTCTGTGTGCACAG 1668
 Db 773 LysProThr-----ProLysAlaAlaSerSerSerSerSerSerSerSerSerSer 791
 QY 1669 AGACAGAGCTCCAGAAAGAAAGACACACAGT-----AAGAAAGGAGTGGTGCCT 1722
 Db 792 AspValSerLysAlaLysLysThrAsnThrAlaValSerLysSerProValThrPro 811
 QY 1723 GCGCGAGTGAAGCTCGGAGAGACAGCCTTGTGAGACAGC 1767
 Db 812 LysAlaValProAlaAlaLysLysGluSerSerSerSerSerSerSerSerSer 826

RESULT 12
 JW0057
 C:Species: Homo sapiens (man)
 C:Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
 C:Accession: JW0057
 R:Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
 J. Biochem. 123, 1119-1126, 1998
 A:Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial cells
 A:Reference number: JW0057; MUID:98269042; PMID:9604001
 A:Accession: JW0057
 A:Molecule type: mRNA
 A:Residues: 1-1684 <SAT>
 A:Cross-references: DBJ:AB003476; NID:g2081606; PIDN:BA119927.1; PID:d1020716; PID:g2081606
 C:Comment: This protein regulates cell growth.
 F:433-439/Region: nuclear location signal
 F:522-527/Region: nuclear location signal
 F:591-596/Region: nuclear location signal
 F:671-676/Region: nuclear location signal

Alignment Scores:
 Fred. No.: 0.00355 Length: 1684
 Score: 169.50 Matches: 139
 Percent Similarity: 30.58% Conservative: 76
 Best Local Similarity: 19.77% Mismatches: 221
 Query Match: 3.61% Indels: 267
 DB: 2 Gaps: 30

US-09-787-016A-1 (1-2610) x JW0057 (1-1684)

QY 226 TCAGGAAAGTGTCCCAAGCTTTAGTGTGGACAGATATGGACAGCAAGGCGACCCG 285
 Db 160 SerGlyGlnAlaValGluGlu-----CysLysGluGluGluGlu 173
 QY 286 AGCAATAGAGGAGCACT-----AAGCCATCAAAACCCACAGC----- 324
 Db 174 LysGlnGluLysGluProSerLysSerAlaGluSerProThrSerProValThrSerGlu 193
 QY 325 -----AAGAGTTTCAGGAAACATGG-----GGTTTTCGAGGACCACT 363
 Db 194 ThrGlySerThrPheLysPhePheThrGlnGlyTrpAlaGlyTrpArgLysLysThr 213
 QY 363 ----- 363
 Db 214 SerPheAspLysProLysGluValGluAlaSerGluLysLysGluGlnGlu 233
 QY 364 -----ATCGCAAGCGAGAGGCGCA 384
 Db 234 ProGluLysValAspThrGluGluAspGlyLysAlaGluValAlaSerGluLysLysThr 253
 QY 385 GGGGACGGGAGGCTGACCCACTGGAGCGGCA----- 417
 Db 254 AlaSerGluGlnAlaHisProGlnGluProAlaGluSerAlaHisGluProArgLysSer 273
 QY 418 -----CCCCACAGCAGCAGCTGGGCGCTGTCCCTCGCGGCGC 453
 Db 274 AlaGluTyrglyLysValGluValGluLeuProSerGluGluGlnVal----- 287
 QY 454 AGTGGGAGGCGAG-----CCCAAGCGCACTGAGCGGCTGGAG 489
 Db 288 SerGlySerGlnGlyProSerGluGluLysProAlaProLeuAlaThrGluValPheAsp 307
 QY 490 CAGTTCCTGACCATTCGCGCGCGCGCGAGGAGCATGCTGCTCTCCCTGGAGCAT 549
 Db 308 GlnLysIleGluValHisGlnGluValAlaGluValHisValSerThrValGlu 327
 QY 550 TCTGTGAGCCACCATCTCTCCCGCCACAGAGCC-----GAGACAGCCTCCGAGCGCAGC 606
 Db 328 -----GluArgThrGluGluGlnLysThrGluValGluGluThrAlaGlySerValPro 345
 QY 607 GTGGAAAGCGCTTCGACACAGAGCGCGCCCGCTGCTGCTCCACAGCTGTGAAGAA 666
 Db 346 AlaGluGluLeuValGluMetAspAlaGluProGlnGluAlaGluProAla-----LysGlu 364
 QY 667 CGACCAAGCTCTTCTGAAAGGTGAAGAGGAGGATGACACAGCATGCCATAGT 726
 Db 365 LeuValLysLysLysGluThrCysValSerGlyGluAspProThrGlnGlyAlaAspLeu 384
 QY 727 GACAGCGATGGCTGACCTTGAAGAG----- 753
 Db 385 SerProAspGluLysValLeuSerLysProGluGlyValValSerGluValGluMet 404
 QY 754 CTTCAAGATCGCTTCGAGGAAAGCGGAAACAGAGAGCC-----ACTGAG 798
 Db 405 LeuSerSerGlnGluArgMetLysValGlnGlySerProLeuLysLysLysLysPheThrSer 424
 QY 799 AGGCCCTTGAAGGATTCAGATGCTGCGGAGAGAGCGCGGAGGAGGCT----- 852
 Db 425 ThrGlyLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 444
 QY 853 -----CCCCCGAGACTGTGGGCTCGAGGCGCAGTGCACACT 888
 Db 445 GluSerGlyGluHisThrGlnValProAlaAspSerProAspSerGlnGluGlnLys 464
 QY 889 GTGGAGGCGCTCTGCCCGCAGTAAGCAGGAGCGCCAG----- 924
 Db 465 GlyGluSerSerAlaSerSerProGluGluProGluLeuThrCysLeuGluLysGly 484
 QY 925 -----AACGATCAGCGGGTGTGTGCCAGGCTGGGAAAGATGACAGAGAGT 972
 Db 485 LeuAlaGluValGlnGlnAspGlyGluAlaGluGluGlyAlaThrSerAspGlyGluLys 504

973 AAGTTGGAGGA----- 984
 505 LysArgGluGlyValThrProTrpAlaSerPheLysLysMetValThrProLysLysArg 524
 984 ----- 984
 525 ValArgProSerGluSerAspLysGluAspGluLeuAspLysValLysSerAlaThr 544
 985 -----AAGCGCGCTCAGGACATCAAGATGAGGAGCGCTGAGACTTGGCC 1029
 545 LeuSerSerThrGluSerThrAlaSerGluMetGlnGluMetLysGlySerValGlu 564
 1030 CGACCGAGCTGAATGTCAGGGTTACGACCCCAAGCCCTGTATTGCACTTTCGCCAG 1089
 565 GluProLysProGlu-----Glu 570
 1090 CTTCAACAACAAGAGTTTATGATTCTGTGACCGCTGCAAGAAATGTTTCATGCGAT 1149
 571 ProLysArgLys-----ValAspThrSerValSerTrpGluAlaLeuLeu 585
 1150 TGTGTGGGCAATTTCTAGGCTCGAGGAGGCTTTTGAAGGATGGAAGACTATATC 1209
 586 CysValGlySerSerLysLysArgAla-----ArgArgGlySerSer----- 599
 1210 TGCCCAAACTGCACCATTTCTGCAAGTGCAGGATGAGACTCAATTCAGAAACGGCA----- 1263
 600 -----SerAspGluGluGlyProLysAlaMetGly 610
 1264 ---GATCAGCAGGAAGCTAAATGAGACCTGGAGATGCTGAT---GGCACCGATTGTACA 1317
 611 GlyAspHisGlnLysAlaAspGluAlaGlyLysAspLysGluThrGlyThrAspGlyLe 630
 1318 AGTATAGGACAAATAGCAGAGAGTCTAGCGAAGACCAAGGATTAAGGTAGTAATGAG 1377
 631 LeuAlaGlySer-----GlnGluHisAspProGlyGlnGly---SerSerSerProGlu 647
 1378 AAAGCTGCAAAATCAAGTGGCAAGAGAAACTCAAGATCTTCCAGCTGTGTAGAGCGG 1437
 648 GlnAlaGlySerProThr-----GluGly 655
 1438 CTGTGTGCTCAAAATGATTGTCGCGCGGTGCTGTACGTGCGCAGCCGACTCGGTG 1497
 656 GluGlyValSer----- 659
 1498 TACTGCAAGTAATGACTGTATCTCTCAACACGCGCCGACGACAATGAAGTTTCTAAGTCA 1557
 660 -----ThrTrpGluSerPheLysArgLeuValThr 669
 1558 GGTAAAGACAGAGCCAAAGCCCTAAAGAAAGATGAGATGAGCCAGAGAGCCCGAGT 1617
 670 -----ProArgLysSerLysSerLysLeuGluGluLysSer 682
 1618 CTTCCGAAATGCGGTGCTCAGCGAGGTATTAAATCTCTTCTGTGCAACAGAGACCAAGCT 1677
 683 GluAspSerLe-----AlaGlySerGlyValGluHisSerThrProAspThrGluProGly 701
 1678 CCAGAAATAAAGAGACCAAGTGAAGAGGACGTGTGTGCTCGCGCGGAGTGAAGCA 1737
 702 LysGluGluSerTrpValSerLeileLysLysPheLeProGlyArgLysLysArgPro 721
 1738 CTCGGGAG-----GAAGCAGCTTGTGAGCAGCAGCAGCGCGTGTGCGGAGGATCAC 1791
 722 AspGlyLysGlnGluGlnAlaProValGluAspAlaGlyProThrGlyAlaAsnGluAsp 741
 1792 AAT-----TACATGCAAGTAAGCCAGAAAG 1818
 742 AspSerAspValProAlaValValProLeuSerGluTyraAspAlaValGluArgGluLys 761
 1819 ACTGCTGCT 1827
 762 MetGluAla 764

RESULT 13

568191 triadin - human
 C:Species: Homo sapiens (man)
 C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 11-Jan-2000
 R:Accession: S68191
 R:Task: N.I.; Eyre, H.J.; O'Brien, R.O.; Sutherland, G.R.; Denborough, M.A.; Foster, P.
 R:J. Biochem. 233, 258-265, 1995
 A:Title: Molecular cloning of the cDNA encoding human skeletal muscle triadin and its lo
 A:Reference number: S68191; MUID:96061957; PMID:7598753
 A:Accession: S68191
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-729 <TAS>
 A:Cross-references: EMBL:U18985; NID:9882222; PIDN:AAA75315.1; PID:9882223
 C:Genetics:
 A:Map position: 6q22-6q23
 C:Superfamily: histone H1

Alignment Scores:
 Pred. No.: 0.00345 Length: 729
 Score: 169.00 Matches: 101
 Percent Similarity: 35.11% Conservative: 83
 Best Local Similarity: 19.27% Mismatches: 168
 Query Match: 3.60% Indels: 172
 DB: 22 Gaps: 22

US-09-787-016A-1 (1-2610) x S68191 (1-729)

QY 532 CTGTCTCCTCGAGGATTTCTGTGAGCCACGCTCTGCCCGCCACAGCCGAGACA 591
 Db 293 ProLeuProThrGluGlnAlaSerArgProThr-----ProAlaSerProAlaLeuGlu 310
 QY 592 GCTCTCGAGGCGAGCGTGGAAAGCGCT-----TCTGAGACC----- 527
 Db 311 GluLysGluGlyGluLysLysLysAlaGluLysLysValThrSerGluThrLysLysLys 330
 QY 628 -----AGAGCGGCCCGCCAGTCTGCCACAGCTGTCTCCACAGCTGTGAAGAA 666
 Db 331 GluLysGluAspLeuLysLysLysSerGluLysGluThrAlaLeuAspValGluLysLys 350
 QY 667 CGACAGACGCTCTTCTCAAAAGGTGAAGAGGAG----- 699
 Db 351 GluProGlyLysAlaSerGluThrLysGlnGlyThrValLysLeuAlaAlaGlnAla 370
 QY 700 -----GATGACCAACATGACACCTCGGATAGTACAGCGATGCGCTTGAA 750
 Db 371 AlaLysLysAspGluLysLysGluAspSerLysLysThrLysLysProAlaGluValGlu 390
 QY 751 GAGCTTCAAGATCGCTTCGCGAGAGCGGAGAACAG-----GAGCCCACTGAGAGGCC 804
 Db 391 GlnProLysGlyLysLysGlnGluLysLysGluLysHisValGluProAlaLysSerPro 410
 QY 805 CTGAAA-----GGGATCCAGATGCTCCCTCGGAGAGAGCGCGGAGGAGGTCGCCGCC 858
 Db 411 LysLysGluHisSerValProSerAspLysGlnValLysAlaLysThrGluArgAlaLys 430
 QY 859 GAGACTGTGGCTCCGAGGCGAGTACACTGTGTGAGCGGCTCTGCGCCAGTAAAGAGGAG 918
 Db 431 GluGluLeuLeuValValSerSer-----LysLysAlaValProGlyLysLysGlu 447
 QY 919 CCGAGACACATCAGGGGCTGTGTCCAGCGCTGGGAAAGATGACAGAGAGTAAGTTG 978
 Db 448 GluLysThrThrLysThrValGluGlnGluLeuLeuArgLysGluLys----- 462
 QY 979 GAGGGAAGCGGCTCAGGACATCAAGATGAGAG----- 1014
 Db 463 SerGlyLysThrSerSerLeuLysAspLysGluProLeuLysGlyLysGluLys 482
 QY 1015 ---CCTGGAGACTTGGCGGACCGAAGCTGAAATGTGAGGTTTACGACCCCAACGCCCTG 1071
 Db 483 ValProAlaSerLeuLysGluLysGluProGluThrLys----- 495
 QY 1072 TATTGCAATTGCGCGCCAGCTCACACACAGGTTTATGATTGCTGTGACCGCTGTGAA 1131

Db	578	oLysSerProAlaThrValLysSerProGlyGluAlaLysSerProSerGluAlaLysSe	598
Qy	732	CGATGCGCTGACCTTGAAAGAGCTTCAGATCGCCTCG--CAGAAGCGGGAGCAGGA	788
Db	598	rProAla--GluAlaLysSerProAlaGluAlaLysSerProAlaGluAlaLysSerPro	617
Qy	789	GCCCACTGAGAGGCCCTCGAAAGGATCCAGAGTCGCTCGGAA-----	833
Db	618	AlaGluAlaLysSerPro--AlaGluAlaLysSerProAlaGluAlaLysSerProAla	636
Qy	834	-----GAACGCGCGGAGAG-----GGTCCCGCCGAGAC	863
Db	637	ThrValLysSerProGly-GluAlaLysSerProSerGluAlaLysSerProAlaGluAl	656
Qy	864	TGTGGGC--TCCGAGCCAGTCACATGTGGAGGGCGCTCGTCCGC-----	906
Db	656	aLysSerProAlaGluAlaLysSerProAlaGluAlaLysSerProAlaGluValLysSe	676
Qy	907	-----AGTAAGCAGAGCCGAGACGATCAGGGGTTGTGCCAGGC	950
Db	676	rProGlyGluAlaLysSerProAlaGluProLysSerProAlaGluAlaLysSerProAl	696
Qy	951	TGGGAAAGATGACAGAGAGTAGTAAATTGGAGGGAAGCGCGCTCAGACATCAAAGATGA	1010
Db	696	a-----GluValLysSerProAlaGluAlaLysSerProAlaGluValLys-----	711
Qy	1011	CGAGCCTGAGACTTGGGCGCAGCG-----AGCCTGAATGTAGGGTTACGACCC	1061
Db	712	-SerProGlyGluAlaLysSerProAlaAlaValLysSerProAlaGluAlaLysSerPr	731
Qy	1062	CAAGCCCTCTATTGCCATTTCGCCCGCAGCTCACACACAGGTTTATGTTGCTGTGA	1121
Db	731	oAlaAlaVal-----LysSerPro-----	737
Qy	1122	CCGCTGTGAAGAAATGGTTTCATGCGGATTTGTGGCATTTCTGAGCTCGAGGAGGCT	1181
Db	738	-----GlyGluAlaLysSerProGlyGluAlaLysSerProAl	750
Qy	1182	TTTGGAAAGAAATGGGAAGACTATATCTGCCAAACTGCACCATTTCTGCAAGTCGAGGA	1241
Db	750	aGluAlaLysSerProAlaGluAlaLysSerPro-----flegluValLysSe	766
Qy	1242	TGAGACTCATTCAGAAACGGCAGCATCAGCAGGAAGCTAAATGGAGACTTCGAGATGCTGA	1301
Db	766	rProGluLysAlaLysThrProValLysGluGlyAlaLys--SerProAlaGluAlaLyl	785
Qy	1302	TGGCACCAGATTGTACAAGTATAGAACAAATAGACCAG-----AGTCTAG	1346
Db	785	sSerProGluLysAlaLysSerProValLysGluAspIleLysProProAlaGluAlaLyl	805
Qy	1347	CGAAGACCAAGGATAAAGGTGAATTTGAGAAGCTGCAATCCAACTGCCAAG--AA	1403
Db	805	sSerProGluLysAlaLysSerProValLysGluGlyAlaLysProProGluLysAlaLyl	825
Qy	1404	GAAACTCAGATCTTCAGCCTGTGATAGAGCGCCCTGGTGCCTCAAAATGTATTGGCCC	1463
Db	825	sProLeuAspValLysSerProGluAlaGlnThrProValGlnGluAlaThrValPr	845
Qy	1464	CGGTGTGTCTCAGTGGCGCAGCCGACTCGGTACTGCG-----	1503
Db	845	o-----ThrAspIleArgProProGluGlnValLysSerProAlaLysGluLysAlaLyl	863
Qy	1504	-AGTAATGACTGTATCTCAAACACGCCGAGACAAATGAAGTTTCTTAAGCTCAGGTAA	1562
Db	863	sSerProGlu-----LysGluGluAlaLysThrSerGluLysValaLProLysLyl	880
Qy	1563	AGAACAAAGCCAAGCT--AAAGAAAGATCAAGATAG--CCAGAGAGCC	1613
Db	880	sGluGluValLysSerProValLysGluGluValLysAlaLysGluProProLysLysVa	900
Qy	1614	CAGT-----CTTCGCAATGCGTGTCCAGCAGGATTTAAATTCCT	1655

Db 900 lGIuGIuGIuThrLeuProThrProLysThrGIuAlaLysGIuSerLysLysAspGI 920
 QY 1656 TTCTGTGTCACAGACACCAAGCTCCA-----GAAAAAAGAGACACCACTGAAGAA 1706
 Db 920 uAlaProLysGIuAlaProLysProLysValGIuGIuLysLysGIuThrProThrGIuLy 940
 QY 1707 GGCA-----GTGGTGTCTCCTCGCGGAGTGAAGCACTCGG-----AAGGAAGC 1751
 Db 940 sProLysAspSerThrAlaGIuAlaLysLysGIuGIuAlaGIuLysLysLysAlaVa 960
 QY 1752 AGCTTGTGAGACAGCAGCCGCTCG 1776
 Db 960 lAlaSerGIuGIuThrProAla 968

RESULT 15
 T42727
 proliferation potential-related protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Sep-2000
 C:Accession: T42727
 R:Witte, M.N.; Scott, R.E. submitted to the EMBL Data Library, November 1998
 A:Reference number: Z22246
 A:Accession: T42727
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1560 ->IT>
 A:Cross-references: EMBL:U03913; NID:G3858884; PID:G3858885; PIDN:AAC72432.1
 A:Experimental source: strain Balb/C
 C:Genetics:
 A:Gene: P2P-R
 C:Function:
 A:Description: involved in hnRNP association and Rb1 binding
 C:Superfamily: RING finger homology
 F:57-107/Domain: RING finger homology <RPN>

Alignment Scores:	
Pred. No.:	0.00512
Score:	167.00
Percent Similarity:	33.43%
Best Local Similarity:	20.06%
Query Match:	3.56%
DB:	2
	Length: 1560
	Matches: 132
	Conservative: 88
	Mismatches: 229
	Indels: 209
	Gaps: 26

US-09-787-016A-1 (1-2610) x T42727 (1-1560)

QY 262 GGTATGACGACAAAGCGACCCGAGCAATGAGAGGACCTAAGGCGCATCAAAACCCACC 321
 Db 811 GlyProGlnGlnLysValAspGlyAspArgGIuLysSerPro-Arg--SerGluProPr 829
 QY 322 AGCAAGAGTTTCAGGAAAACATGGGTTTTCGAAAGACACACTATCGCCAGCGAGAGCGG 381
 Db 829 oLeuLysLysAlaLysGIuGIuAlaThrLysIleAspSerValLysProSerSerSe 849
 QY 382 GCAGGGAGCGGAGGCTGACCCACTGGAGCCGCCACCCACAGCAGACGACTGGG--- 437
 Db 849 rGlnLysAspGlnLysValThrGlyThrProArgLysAlaHisSerLysSer-AlaLysG 869
 QY 437 ----- 437
 Db 869 LuHisGlnGIuAlaLysProAlaLysAspGIuLysValLysLysAspCysSerLysAspI 889
 QY 438 -----CTGTCTCTCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 489
 Db 889 lElysSerGIuLysProAlaSerLysAspGIuLysAla-LysLysProGIuLysAsnLys 908
 QY 490 CAGTTCTGTACCATTTGGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 549
 Db 909 LeuLeuAspSerLysGIuLysArgLysArgLysThrGIuGIuLysSerValAspLys 928
 QY 550 TCTGTGTAGCCCGAGTCTCTCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 597
 Db 929 AspPheGlnSerSerMetLysIleSerLysValGlnGlyThrGlnLysValLysPro 948

```
QY 598 -----GAGGCGAGCGTGAAGCGCTTCTGAGACGAGAGCGGCC 639
Db 949 SerProLysArgLysMetGluGlyAspValGluLysLeuGluArgThr-----Pro 965
QY 640 CAGTCGTCTCCAGAGCTGTGAAGAGACGACGAGCTCTTCTGAAGAGTGAAGAGGG 699
Db 966 GluLysAspLysIleAlaSerThrThrProAla-----LysLysIleLysLeuAsn 983
QY 700 GATGACACGATGACACCTCCGATAGTACACGATGCGCTGACCTTGAAGAGCTTCAG 759
Db 984 ArgGlu-----ThrGlyLysLysIleGly 991
QY 760 AATCGCCTTCGAGAGCGGGAACAGAGGCCACTGAGAGGCCCTCAAGAGGATCCAG 819
Db 992 AsnAlaGluAsnAlaSerThrThrLysGluProSerGluLys---LeuGluSerThrSer 1010
QY 820 AGTCGCGCTG-----CGAAGAAGCGCGGAGGAG 849
Db 1011 SerLysIleLysGlnGluLysValLysGlyLysAlaLysArgLysValAlaGlySerGlu 1030
QY 850 GGTCCGCGCGAGACTGTG----- 867
Db 1031 GlySerSerThrLeuValAspTyrThrSerThrSerThrGlyLysProVal 1050
QY 868 -----GGTCCGAGCCAGTCACACTGTGAGGCGCTCTCCAGTAAGCAGAGGCC 921
Db 1051 ArgLysSerGluGluLysThrAspThrThrValLysThrMetGluGluTyr 1070
QY 922 GAGAACGAT-----CAGGCGGTGTG-----TCC 945
Db 1071 AsnAsnAspThrAlaProAlaGluAspValIleMetIleHisValProGlnSer 1090
QY 946 CAGGCTGGGAAGATGACAGAGAGTAAGTTGGAGGGAAGCGGCTCAGCACATCAA 1005
Db 1091 LysTrpAspLysAspPheGluSerGluGluGlu-----AspValLys 1105
QY 1006 GATGAGGAGCCT---GGAGACTTGGCCGACCG----- 1035
Db 1106 ThrThrGlnProIleGlnSerValGlyLysProSerIleLysAsnValThrThr 1125
QY 1036 AGCCT-----GAATGTAGGTTAGACCCCAACGCCCTGTAT 1074
Db 1126 LysProSerAlaThrAlaLysTyrThrGluLysGluSerGluGlnProGluLysLeuGln 1145
QY 1075 TGCAATTGCGCCAGCCTCACAAACAGGTTTATGATTGCTGTGACCGCTGTGAAGAA 1134
Db 1146 LysLeuProLysGluAlaSerHis-----GluLeuMetGlnHis 1158
QY 1135 TGCTTTTCATGCGGATTCGTGGGCATTTCTGAGCTCGAGGAGGCTTTTGAAGG--- 1191
Db 1159 GluLeuArgSerSerLysGlySerAlaSerSerGluLysGlyArgAlaLysAspArgGlu 1178
QY 1192 ---AATGGGAGACTATATCTGCCCAAC----- 1218
Db 1179 HisSerGlySerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLys 1198
QY 1219 ---TGACCACTTCGACGAGGATGAGATCTATTCAGACGACGACATCAGCAGGAA 1275
Db 1199 GluSerThrValAspArgLeuSerGluGlnGlyHisPheLysThrLeuSerGlnSerSer 1218
QY 1276 GCTAAATGAGACCTGGAGATGCTGTGACCGATGTTACAAGTATAGGAACAATAGAG 1335
Db 1219 LysGluThrArg----- 1222
QY 1336 CAGAACTCTAGCGAAGCAAGGATAAGGGTAGAATTGAGAAAGCTGCAAAATCCAAGT 1395
Db 1223 ---ThrSerGluLysHisGluSerValArgGlySerSerAsnLysAspPheThrProGly 1241
QY 1396 GGCAGAGAACTCAAGATCTCCAGCCTGTGATAGAGCGGCTGCTGCCTCAAAATGT 1455
Db 1242 ArgAspLysLysValAsp----- 1247
```

```
QY 1456 ATTGGCCCCGGTGTCTGTACGTGGCGCAGCCGACTCGGTACTGCAATGACTGT 1515
Db 1248 -----TyrAspSerArgAspTyr 1253
QY 1516 ATCTCTAAACACGCGCGCAGCACAATGAGTTTCTAAGCTCAGGTAAAGAACAGAGCCA 1575
Db 1254 SerSerSerLysArgArgAspGluArgGlyGluLeuAlaArgArgLysAspSerProPro 1273
QY 1576 AGCCTAAAGAA-----AAGATGAAGATGAAGCCAGAGAGAGCCAGTCTT 1620
Db 1274 ArgGlyLysGluSerLeuSerGlyGlnLysSerLysLeuArgGluGluArg---AspLeu 1292
QY 1621 CCGAAATGCGGTGCTCAGGCGAGGTATTAATCTCTCTGTG----- 1662
Db 1293 ProLysLysGlyAlaGluSerLysSerAsnSerSerProProArgAspLysLysPro 1312
QY 1663 -----CACAGAGACAGCTCCAGAAAAGAGAGACACAGTGAAGAGCGCATG--- 1713
Db 1313 HisAspHisLysAlaProTyrGluThrLysArgProCysGluGluThrLysProValAsp 1332
QY 1714 -----GTGTCCTCCTGCGCGAGT-----GAAGCA 1737
Db 1333 LysIleSerGlyLysGluArgGluLysHisAlaAlaGluAlaArgAsnGlyLysAspSer 1352
QY 1738 CTCGGGAAGGAAGCAGCTTGTGAGAGCAGCAGCGCTGCTGGCGGCGGATCACAATTAC 1797
Db 1353 SerGlyGlyLysLeuProCysIleLeuAsnProProAspLeuProMetGluLysGluLeu 1372
QY 1798 AATCAGTAAAGCCAGAGAAAGACTGCTGCT---CCCTCGCCCTCAGCTG 1842
Db 1373 AlaValGlyGlnValGluLysSerAlaValLysProLysProGlnLeu 1388
```

Search completed: April 28, 2004, 11:01:34
Job time : 86 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2004, 11:00:28 ; Search time 115.5 Seconds

(without alignments)
12527.529 Million cell updates/sec

Title: US-09-787-016A-1

Perfect score: 4689
Sequence: 1 ctccgtggcgcgcgcac.....actcttaagatcatatctcg 2610

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 2276240

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-DB=/cgn2_1/USPTO_spool_p/US09787016/runat_28042004_094331_27112/app_query.fasta_1.2759
-B=Published Applications AA -OFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNIT5-bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09787016 @CGN 1 1 17 @runat_28042004_094331_27112
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCKS=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
------------	-------	-------	--------	-------	-------------

1	958	20.4	181	9	US-09-925-301-1169
2	181.5	3.9	268	9	US-09-764-864-1258
3	180.5	3.8	669	10	US-09-823-187-88
4	180.5	3.8	669	10	US-09-863-776-53
5	180	3.8	19595	15	US-10-084-846A-3
6	177.5	3.8	19582	15	US-10-084-846A-7
7	176	3.8	3551	12	US-10-263-929-144
8	176	3.7	19725	15	US-10-084-846A-4
9	172	3.7	676	10	US-09-823-187-24
10	172	3.7	676	10	US-09-863-776-14
11	171.5	3.7	555	14	US-10-106-698-5847
12	171	3.6	678	14	US-10-203-708-28
13	171	3.6	678	10	US-09-823-187-87
14	171	3.6	678	10	US-09-863-776-52
15	171	3.6	717	14	US-10-203-708-29
16	171	3.6	3262	15	US-10-379-381-4
17	170.5	3.6	19695	15	US-10-084-846A-3
18	170	3.6	667	10	US-09-863-776-16
19	170	3.6	667	10	US-09-863-776-18
20	169.5	3.6	19725	15	US-10-084-846A-4
21	169	3.6	1404	9	US-09-811-045A-1
22	169	3.6	2781	12	US-10-263-929-122
23	169	3.6	19723	15	US-10-084-846A-5
24	168.5	3.6	518	9	US-09-925-299-862
25	168.5	3.6	518	10	US-09-925-299-862
26	168	3.6	670	10	US-09-823-187-86
27	168	3.6	670	10	US-09-863-776-51
28	168	3.6	671	10	US-09-946-374-308
29	168	3.6	671	10	US-09-823-187-85
30	168	3.6	671	10	US-09-863-776-20
31	168	3.6	671	12	US-10-206-915-346
32	168	3.6	671	12	US-10-199-670-346
33	168	3.6	671	12	US-10-201-858-346
34	168	3.6	671	12	US-10-205-890-346
35	168	3.6	671	12	US-10-208-024-346
36	168	3.6	671	12	US-10-201-853-346
37	168	3.6	671	12	US-10-174-581-346
38	168	3.6	671	12	US-10-176-483-346
39	168	3.6	671	12	US-10-176-749-346
40	168	3.6	671	12	US-10-176-914-346
41	168	3.6	671	12	US-10-176-915-346
42	168	3.6	671	12	US-10-006-485A-308
43	168	3.6	671	12	US-10-013-907A-308
44	168	3.6	671	12	US-10-015-499A-308
45	168	3.6	671	12	US-10-176-484-346

ALIGNMENTS

RESULT 1

US-09-925-301-1169
; Sequence 1169, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P106
; CURRENT APPLICATION NUMBER: US/09925301
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1169
; LENGTH: 181
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-301-1169

Alignment Scores: 5.29e-58 Length: 181
Pred. No.: 181

Score:	958.00	Matches:	179
Percent Similarity:	98.35%	Conservative:	0
Best Local Similarity:	98.35%	Mismatches:	2
Query Match:	20.43%	Indels:	2
DB:	9	Gaps:	0

US-09-787-016A-1 (1-2610) x US-09-925-301-1169 (1-181)

998	QY	ACATCAAGAATGAGGAGCGTGGAGACTTG- GGCGGACCGAGCGCTGAATGTGAGGGTAC	1056
	Db	1 ThrSerLysMetArgSerLeuGluThrLeuGlyArgProLysProGluCysGluGlyTyr	20
1057	QY	GACCCCAACGCCCTGTATTGCATTGGCCGACGCTCACAACAACAGGTTTATGATTGC	1116
	Db	21 AspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnArgPheMetIleCys	40
1117	QY	TGTGACCCCTGTGAAGATGGTTTCATGGCGCATGTGTGGCAATTCTGAGCGCTCGAGGG	1176
	Db	41 CysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgGly	60
1177	QY	AGGCTTTTGGAAAGGAATGGGAGACATATATCTGCCCAACTGCACCATTTCTGCAAGTG	1236
	Db	61 ArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnVal	80
1237	QY	CAGGATGAGCTCATTCAGAAAGCGCATGACGAGAGCTTAATGAGAGCTCGGAGAT	1296
	Db	81 GlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaGlyTyrArgProGlyAsp	100
1297	QY	GCTGATGCCACCGGATTGTACAAGTATAGGAAACAATAGACAGAAGTCTAGCGAAGACAA	1356
	Db	101 AlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGln	120
1357	QY	GGGATAAGGCTGCAATGAGAAGCTGCAATCCAAAGTCGCAAGAGAAACTCAAGTC	1416
	Db	121 GlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLeuLysIle	140
1417	QY	TTCCAGCTGTGTAGAGGCGCTGTGGCTCAAAATGTATTGGCCCGGCTGCTGCAC	1476
	Db	141 PheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysHis	160
1477	QY	GTGGCGGACCGGACTGGGTGTAAGTCAAGTAAGTGTATCCTCAACACGCGCGGACG	1536
	Db	161 ValAlaHis-ProThrArgCysThrAlaValMetThrValSerSerAsnThrProGlnArg	180
1537	QY	ACAA 1540	
	Db	180 gGln 181	

RESULT 2

US-09-764-864-1258
; Sequence 1258, Application US/09764864
; Patent No. US20020132753A1

GENERAL INFORMATION:

; APPLICANT: Rosen et al.

REICMAN, ROSEN ET AL.; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT223

FIELD REFERENCE: F1223
CURRENT APPLICATION NUMBER: US/09/764,864

CURRENT FILING DATE: 2001-01-17

```

; COMMENT: PENDING DATA: 2001-01-01
; Prior application data removed - consult PALM or file wrapper
;

```

; NUMBER OF SEQ ID NOS: 1792

; SOFTWARE: Pat

COL INDEX	SEQ ID NO	1258
1	1	1258

; LENGTH: 26

```

;
; TYPE: PRT
;

```

ORGANISM:

```

; CREATE TABLE, FROM
; FEATURE:

```

NAME/KEY: SITE

```

/ NAME/LOC: 0112
; LOCATION: {162}

```

OTHER INFORMATION:

NAME/KEY: SITE

; LOCATION: (209)

OTHER INFORMATION:

```

; LOCATION: (264)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (265)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (268)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1258

```

Alignment Scores:

Alignment Scores:		
Pred. No.:	0.000459	Length: 268
Score:	181.50	Matches: 52
Percent Similarity:	43.0%	Conservative: 29
Best Local Similarity:	27.6%	Mismatches: 52
Query Match:	3.87%	Indels: 45
DB:	9	Gaps: 10

UIS-09-787-016A-3 (1-2610) x UIS-09-764-864-1258 (1-268)

```

934 GGGGTTGTGTCCTCCAG---GCTGGGAAGAATGACAGAGAGAGTAGTAAGTTGAGGGAAGCGC 990
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
29 GlyValValArgGluCysAlaGlySerGlyAspMetGluGly----- 42
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
991 GCTCAGGACATCAAGAATAGGAGCGCTGGAGACTGTTGGCGCGACCGAGCCTGAATGTCAG 1050
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
43 -----AspGlySerAspProGluProProAspAlaGlyGluAspSer---LysSerGlu 59
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1051 GATTACGACCCCAACGCCCTGTATTGCATTTGCCGCCAGCCTCACAAACAGGTTTATG 1100
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
60 AsnGlyGluAsnAlaProIleTyrCysIleCysArgLysProAspIleAsnCysPheMet 79
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1111 ATTTGCTGTGACCGCTGTGAAGAATGTTTCATGGCGATTGTGGGCATTTCTCAGGCT 1170
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 IleglyCysAspAsnCysAsnGluTrpPheHisGlyAspCysIleArgIleThrGluLys 99
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1171 CGAGGAGGCTTTTGGAAAGGAAAGGAGACTATATCTGCCCAACTGC----- 1220
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
100 MetAlaLysAlaIle-----ArgGluTrpTyrCysArgGluCysArgGluLys 115
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1222 ---ACCATTCCTGCAAGTGCAG-----GAT 1242
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 AspProLysLeuGluIleA-GTyrArgHisLysLysSerArgGluArgAspGlyAsnGlu 135
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1243 GAGACTCATTCAGAAACGGCAGACTACGAGGAGCTAAATGGAGACCT---GGAGATGCT 1299
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 ArgAspSerSerGluProArgAspGluGlyGlyArgLysArgProValProAspPro 155
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1300 GATGGCCCGGATTCTACAGTATAGGAACAATAGACAGACAAGCTCTAGCGAAGACCAAGGG 1355
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156 Asp-----LeuGlnArgArgAla**SerGlyThrGly 166
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1360 ATAAGGGTAGAATTGAGAAA---GCTGCAAATCTCAAGTGGCAGAGAGAACTCAAGATC 1416
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 ValGlyAlaMetLeuAlaArgGlySerAlaSerProHisLysSerSerPro----- 183
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1417 TTCAGGCTGTGATAGAGCGGCT 1440
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
184 ---GlnProLeuValAlaThrPro 190
QY      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-09-823-187-88
; Sequence 88, Application US/09823187
; Publication No. US20030096952A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Liu, Xiaohong
; APPLICANT: Majumder, Kumud
; APPLICANT: Padigara, Muralidhar
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Spaderna, Steven K

```

Alignment Scores:					
Pred. No.:	0.000683	Length:	669		
Score:	180.50	Matches:	131		
Percent Similarity:	33.44%	Conservative:	79		
Best Local Similarity:	20.86%	Mismatches:	193		
Query Match:	3.85%	Indels:	225		
DB:	10	Gaps:	29		
 US-09-787-016A-1 (1-2610) x US-09-823-187-88 (1-669)					
QY	220	AGGATTTCAGGGAAAGTGTCCAAGCTTTCAAGTGTGTGGACGAGGTATGACGACAAAGGC	279		
Db	171	ArgAlaSerSerAspLeuaspglnAlaSerValSerProSerGluGluAspSerGlu	190		
QY	280	GACCCGAGCAATGAGAGCGCACCTTAAGGCCATCAAACCACCCACGACAPAGAGTTTCAGAAA	339		
Db	191	SerProSerGluSerGlu-----LysThrSerAspGlnAspPhe-----	203		
QY	340	ACATGGGGTTTTCGNAAGCAGCACACTATCGCCAGCGAGAGGGCGCAGGGGACGCGGAGGCT	399		
Db	204	-----ThrProGluIysLysThrAlaAla-----	211		
QY	400	GACCCACTGGAGCCGCCACCCCCACGACGAGCTGGGGCTCTCCCTCGCGCGCAGTGAG	459		
Db	212	-----ArgProArgArgGlyProLeuGly-----Gly	221		
QY	460	AGGCGACCCAAAGCGCATGAGCGGTGGAGCAGTTCCTGACCATTCGCGCGCGCGCGCGGC	519		
Db	222	ArgLysLysLysValProSerAlaSerAspSerAspSerLysAlaAspSerAspGly	241		
QY	520	AGGAGGACATGCTGTCTCCCTGGAGGATTCGTGTGAGCCCAAGTCTGCCCGCCGCCACA	579		
Db	242	AlaLysGluGluProValVal-----ThrAlaGlnProSer-----ProSerSer	256		
QY	580	GACGCGGAGACAGCTCTCGAGGGCGCAGGTGAAGAAGCGCTTCTGAGACACAGAAGCGCGCCC	639		
Db	257	SerSerSerSerSerSerSerSerSerSerAspValSerValLysLysProPro	276		
QY	640	CAGTCTGCTTCACACAGCTGTGAAG-----	663		

Db 365 GlySer-----SerGlyGluGluLeuGluAsp 373
QY 928 GATCAGGGGTTGTCCAGCTGGGAAGATGACAGAGAGTAAGTTGGAGGAAG 987
Db 374 GluGluProVal-----LysLysArgSerArgLysAlaArgLysArg 387
QY 988 GCGGCTCAGGACATCAAGATCAGAGCTT-----GGAGACTTGGCGCA----- 1032
Db 388 GlyThrProSerSerSerAspSerGluProGluGlyGluLeuGlyLysLysLys 407
QY 1033 -----CCGAAGCTGAATGTAGAGGTACAGCCCAAGCCCTGTATTGCAATTTGCCGC 1086
Db 408 LeuAlaLysLysSerGlnLeuProGlySerGluSer-----AlaArg 421
QY 1087 CAGCCTCACAACAACAGGTTTATGATTCTGTGACCGCTGTGAAGAATGGTTTCATGCC 1146
Db 422 LysPro----- 423
QY 1147 GATTGTGTGGCATTTCTCAGGCTCAGAGAGCTTTTGGAAAGGAATGGGAAGCTAT 1206
Db 424 -----GlyGlnLysGluLysArgGlyArg----- 431
QY 1207 ATCTGCCAAACTGCACCATTTCTGCAAGTGCAGATGAGACTCATTCAGAAACGGCAGAT 1266
Db 432 -----Pro 432
QY 1267 CAGCAGGAAGCTTAATGAGACCTGGAGATGCTGTGCAACCGAATGTACAAGTATAGGA 1326
Db 433 AspGluLysProArgAlaArgProValLysValGluArgThr----- 446
QY 1327 ACAATGAGCAGAGTCTACGAGAACCAAGGATTAAGGTTAGATTGAGAAGCTGCA 1386
Db 447 -----ArgLysArgSerGlyLysLeuSerLeuGluArgLysGlyLysLysLys 463
QY 1387 AATCCAAGTGGCAAGAACTCAAG-----ATCTTCCAGCTGTG 1428
Db 464 GluProSerValGluGluArgLeuGlnLysLeuHisSerGluLeuLysPheAlaLeuLys 483
QY 1429 ATAGAGCGCTGGTGCTCAAAATGATTATGGCCCC-----GGTCTGTCTAC 1476
Db 484 ValAspAsnProAspValArgLysCysLeuSerAlaLeuGluGluLeuGlyThrLeuGln 503
QY 1477 GTGGCG-----CAGCCCGACTCGGTG----- 1497
Db 504 ValThrSerGlnLeuLeuGlnLysAsnThrAspValValAlaThrLeuLysLysLysLeuArg 523
QY 1498 ---TACTGTCAGTAACTGATCTATCTCTCAACACAGCCGCA----- 1533
Db 524 ArgTyrLysAlaAsnLysAspValMetAlaLysAlaAlaGluValTyrThrArgLeuLys 543
QY 1534 -----CCGACATGAAGTTTCTAAGCTCAGCTAAGAA 1566
Db 544 SerArgValLeuGlyProLysValGluAlaLeuGlnLysValAsnLysAlaGlyAlaGlu 563
QY 1567 CAGAAGCAAAAGCTAAAGAAAGATGAAGTGAAGCCAGAGAGCCAGTCTTCGAAA 1626
Db 564 LysGluArgAlaAspAsnGluLysLeuGluGluGlnPro----- 576
QY 1627 TGGGTGCTCAGGAGTATTAAATCTCTCTGTGCAC-----AAGAGACAGCT 1677
Db 577 ---GlyGluGlnAlaProArgGluLeuAlaGluAspGluProSerThrAspArgSerAla 595
QY 1678 CCAGAAAAAAGAGACACACAGTCAAGAGCAGTGGTGTCTGCTGCGAGTGAAGCA 1737
Db 596 ProValAsnGlyGluAlaThrSerGlnLysGlyGluAsnMetGluAspArgAlaGlnGlu 615
QY 1738 CTCGGGAAGGAAGCAGTGTGTGAGAGCAGCAGCCGCTGCTGGGGGAGCGATCAC----- 1791
Db 616 AspGlyGlnAspSer-----GluAspGlyProArgGlyGlySerSerGluGluLeuHis 633
QY 1792 -----AATTACAATGCAGTAAGCA-----GAA 1815
Db 634 AspSerProArgAspAsnSerProAlaLysProGlyAsnGluArgGlnAspHisGlu 653

QY 1816 AAGACTGCTGCTCCTCCCTGCGCGTCA 1839
Db 654 ArgThrArgLeuAlaSerGluSer 661
RESULT 5
US-10-084-846A-3
; Sequence 3, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUELENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 3
; LENGTH: 19695
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-3
Alignment Scores:
Pred. No.: 0.00178 Length: 19695
Score: 180.00 Matches: 158
Percent Similarity: 30.64% Conservative: 77
Best Local Similarity: 20.60% Mismatches: 259
Query Match: 3.84% Indels: 273
DB: 15 Gaps: 33
US-09-787-016A-1 (1-2610) x US-10-084-846A-3 (1-19695)
QY 209 TGGTGTATTATTAGGATTTTCAGGAAAGTGTCCAAAGCTTTCAGTGTGTGAGCAGGTATGG 268
Db 10248 TptTyrPheValAlaAspArgGlyProLeuSerAlaArgSerProArgGlyGluValHis 10267
QY 269 AGCAAAAGCGACCGACCAATGAGGAGGACCTAAGCCATCAACCCACCCAGCAAG 328
Db 10268 SerThrProIle-----GlnLeuProProProSerProSerArg 10281
QY 329 AGT-----TCAGAAAACATGGGGTTCGAAAGGACCACTATCGCCAAAGCAG 376
Db 10282 SerProAsnAlaIleSerGlyArgProAlaCysSerArgSerThrSerProSer--- 10300
QY 377 AGGGCGCAGGGGACGCGG-----AGGCTGACCCACTCGAGC 412
Db 10301 -----GlyGlyArgValCysThrProCysAlaAlaCysAlaProAsnThrSer 10317
QY 413 CCGCACCCCCACAGCAGCAGCTGGGCTGTCTCCCTGGGCGCAGTGGGAGCGAGCCCAAGC 472
Db 10318 ArgCysArgArgArgProAlaSerThrThrProGlyAlaProGlyProCysThrThr 10337
QY 473 GCACCTGAGCGGTGGAGCAGTTCCTGACCATGCGCGCGCGCGCGAGGAGGA----- 526
Db 10338 ProSerGlyArgThrProSerProThrArgProArgSerAlaSerSerThrSer 10357
QY 527 -----GCATGCTGTCTCTCTCGAGG 547
Db 10358 TptThrSerProArgThrProGlyThrArgAlaIleArgCysCysAlaProThrArg 10377
QY 548 ATTCTGTGAGCCCACTGTCTGCGCCCGCCACAGACCGC-----AGACAGCTCGAGG 601
DB: 15 Gaps: 33

[illegible]

```

RESULT 6
US-10-08
; Sequen
; Public
; GENERA
; APPLI
; APPLI
; APPLI
; APPLI
; TITLE
; FILE
; CURRE
; CURRE
; PRIOR
; PRIOR
; PRIOR
; PRIOR
; NUMBE
; SOFTW
; SEQ ID

```

QY	931	CAGGGG	-----TTGTGTC	CCAGCGCTGG	AAAGATG	-----	961								
DB	3573	rTTPGlyHis	gInLeuCys	argGln	alaSerArg	LeuAlaSerThr	LeuThrProCys	3593							
QY	962	---ACAGACAG	AGTAA	TTGG	-----	-----	979								
DB	3593	sArg	ArgSerArg	ValSerTrp	AlaValThr	GlnSerSerGly	ThrAlaThrGln	3613							
QY	980	-----AGCGA	AGCGCGCTC	-----	-----	-----	994								
DB	3613	Yly	SerSerSer	AlaSerThr	ArgArgLeuPhe	CysGluTrp	LeuProProSer	3633							
QY	995	-----	-----	-----	AGCACATCA	AAAGATG	CGGAG	1014							
DB	3633	rAla	AlaThrAla	LeuArgArg	SerGlyGln	AlaValArgThr	SerArgAsnArgGln	3653							
QY	1015	C-----	-----	-----	-----	-----	1015								
DB	3653	yGly	ProCysGly	PheSerGly	ThrHisCys	ThrValGly	ProIleSerMet	ProSer	3673						
QY	1016	-----CTGG	GAGACTTTGG	CGCGACCGA	-----	-----AGCCTG	AA	1044							
DB	3673	rArg	ProGlyTrp	AsnArgAla	TrpThrSer	ArgCysAla	ThrProSerMet	ThrLeu	11	3693					
QY	1045	TGT	GAGGGTTAC	GACCCCA	AGCCCTGT	-----ATTGC	ATTTC	CCGCCAG	CTCAC	1095					
DB	3693	eThr	AsnAla	ThrThrPro	ArgProCys	GlyAsnSerSerGly	TrpSerHis	SerAla	Thr	3713					
QY	1096	AA	CAACACAGG	TTTATG	ATTGCTG	CTGACCGC	-----TGTC	CAAAATGG	TTTCA	1142					
DB	3713	rSer	ArgSerPro	ThrTrp	AlaValThr	ProMet	MetLeu	CysArgPro	SerSerTr	3733					
QY	1143	TGC	CGATTTGTG	GGCA	TTTCTG	AGGCTC	AGGAGG	CTTTTGG	AAAGGA	TGGGA	1202				
DB	3733	pVal	IleCys	SerGly	ValPro	ArgHis	ProGly	Arg-----Ser	ArgArgGly	ValArg	3751				
QY	1203	CT	ATATCTG	CCCAACTG	CCACCATTG	CAAGTGC	AGAGTGC	AGACTCAT	TTCAG	AAACCGC	1262				
DB	3751	gCys	AlaValSer	TrpCys	ArgLeu	LeuAlaSer	ValAsn	GlnArgSer	ThrAla	ProMe	3771				
QY	1263	AG	ATCAGCAG	CAAGTAA	ATG	GGAGACTG	GAGATG	CTGTG	ATG	GCACCA	CTT	1320			
DB	3771	tSer	Phe	ArgGlu	ValPro	TrpPro	ValAsp	SerThrSer	AlaAla	ProThr	SerPr	3791			
QY	1321	AT	AGGACAT	TAG	ACGAGA	AGTCTAC	GACGAC	CAAGGAT	AAAGGT	-----AG	1370				
DB	3791	oThr	GlyThr	Gly	ArgGln	AlaGly	AsnValSer	AlaGly	ArgGln	AlaSer	ThrVal	3811			
QY	1371	AA	TGAAAG	AGCTG	CAAAATCCA	AGTGCC	ACAGAG	AAACTCA	AGATCT	TTC	CCAGC	CTGTG	AT	1430	
DB	3811	gVal	SerMet	LeuThr	ArgPro	SerMet	ArgSer	ArgArg	SerVal	-----	3820	20			
QY	1431	AG	AGGCGCTG	GTGGCTCA	AAATGAT	TGCTG	CCCGCGGT	-----GCTG	TAC	CTGG	-----	1480			
DB	3829	-----Ser	GlySer	GlnTrp	Pro	LeuAla	AlaGly	ThrSer	SerAla	ThrTrp	Thr	3846	46		
QY	1481	-----CG	ACGCGG	ACTCGG	TGTACTG	TCAGTAA	TGCTGTAT	1517							
DB	3846	hr	Met	Leu	ArgLeu	ArgArg	AsnHis	AlaPro	CysSer	AspCys	ArgThr	Ser	3866		
QY	1518	CT	CAAAAC	AGCCG	CAGCAAT	GAAGTTCT	TAAAGCTC	AGGTAA	GAAGAC	AGACCA	1577				
DB	3866	hr	Arg	Val	ArgPro	AlaPro	-----Ala	Leu	Val	Val	AsnSer	ArgArg	3881		
QY	1578	GC	TTAAAG	AAAGAT	GAAGTGA	AGCCAG	CAGACGCC	AGTCTCC	GAAATCG	CGTG	CTCA	1633			
DB	3881	al	Leu	Arg-----	-----Pro	TrpTrp	Met	Arg	Val	Leu	Leu	3892			
QY	1638	GC	AGGTATTA	-----	-----	-----	-----	-----	-----	-----	-----	1650			
DB	3892	rg	ValAla	TrpThr	Val	Gly	AspSerTrp	Met	Arg	Leu	TrpLeu	ProGly	SerIle	Leu	3912

Alignment Scores:	0.00265	Length:	19652
Pred. No.:	177.50	Matches:	165
Score:	31.92%	Conservative:	53
Percent Similarity:	24.16%	Mismatches:	248
Best Local Similarity:	3.79%	Indels:	219
Query Match:	15	Gaps:	32
DB:			
US-09-787-016A-1 (1-2610) x US-10-084-846A-7 (1-19652)			
Qy 284	CGAGCATGAGGAGCGACCTTAAGCCATCAACCCCA	-----CCAGCAAGAGTTCCAGGA	337
Db	3303 ArgAlaProArgArgVal--ProAlaGluProTIPolGlySerAlaAlaGly	3321	
Qy 338	AAACATCGGGTTTCGAAGGACCACTATCCCAAGC	-----GAGAG	379
Db	3322 ArgHisGlyArgSerGlyProArgSerProThrAlaCysThrArgSerGlyArgG	3341	
Qy 380	CGCAGGGGAGCGAGCGGTGACCCACTGAGCGCGCCACCCACAGCAGCAGCACTGGCC	439	
Db	3342 AlaArgSerArgArgThrArgValProCysGlyArgAlaHisThrAlaAlaArgG	3361	
Qy 440	TGTCCTCGCGCGCAGTGGGAGGCGAGCCCAAGGCGCACTGAGCGCGTGAGCAGTCTCTGA	499	
Db	3362 ProProArgGluSerProArgPheProSerAla	3377	
Qy 500	CCATPGCGCGCGCGCGGAGGAGAGCATGCTGTCTCTCTGGAGGATTCGTGTGAGC	559	
Db	3378 ProHisArgGlyArgAlaArg--GlyThrCys--SerProArgArgAlaAspThrArg	3395	
Qy 560	CCAGCTCTCTGCCCGCCACAGACCGC	-----586	
Db	3396 ProArgProAlaGlyCysArgArgProAlaProAlaGlyThrGlyArgCysArgGluAla	3415	
Qy 587	-----AGACAGCCTCCGAGGGGAGCGG	607	
Db	3416 ValAlaGlyThrArgGThrArgTIPolGlySerAlaValHisArgCysArgProArgProAla	3435	
Qy 608	TG-----GAAGCGCTTCGAGACAGAGCGCCCGCAGTCTGCTTCACCA	654	
Db	3436 SerArgAlaArgArgGAlaAlaSerAlaGlyArgSerGlySerProSerAlaSerGly	3455	
Qy 655	GCTGTGAAGGAACACCGACGCTCTTCGAAAGAGTGAAAGCGGGGATGACCCAGTATGAC	714	
Db	3456 ArgGlySerProArgGlyGlyProProSerArgCysProGlySerGlyArgSerAlaAla	3475	
Qy 715	ACCTCCGNTAGTGACGGATG-----GCTGTGACCTTGAAGAGCTTCAGATCGCCTT	768	
Db	3475 aSerProArgAlaArgAlaPheProArgAlaGlyArgAlaArgSerSerArgGlyAlaVa	3495	
Qy 769	CCGAGGAAGCGGGGAACAGGACCCCACTGAGAGCGCCCTGAAAGGGATCCAGATCGCC	826	
Db	3495 lArgTIPAlaGlyArgArgArgProArgArgHisPro-----GlySerProAlaAlaPr	3513	
Qy 827	-----TGCGGAG	834	
Db	3513 oArgSerArgArgArgProArgArgGThrAlaArgArgProHisGlyArgCysProAr	3533	
Qy 835	AAGCGCGGAGGAGGGTCCCGCGAGACTGTGGGCTCCG-----AGGCCA	880	
Db	3533 gSerGlyArgArgArgSerProSerAlaProAlaProSerAlaArgProAlaGlyTy	3553	
Qy 881	----GTGACACTGTGGAGGGGTCC-----TGCCAGTAGACGAGGCGCGAGAACGAT	930	
Db	3553 rGlyValGlyArgCysArgSerSerArgGluCysProAlaAlaArgAspAlaArgSerSe	3573	

QY 1659 TGTGCAAGAGAC-----CAGCTCCAGAAAAAAGAGACCACTGAAGAA 1706
 Db 3912 laCyethrArgGlnCysGlyThrProSerThrSerLeuThrSerMetProArgMetProI 3932
 QY 1707 GGCAGTGGTGGTCCCTCGCGGAGTGAAGCACTCGGGAAGGAAGCAGCTTGTGAGAGCAG 1766
 Db 3932 leSerGlyTrpLeuMetMetGlyTrpLeuHisSerGlyThrGlyProLeuValArgThra 3952
 QY 1767 CACGC 1771
 Db 3952 rgArg 3953
 RESULT 7
 US-10-263-929-144
 ; Sequence 144, Application US/10263929
 ; Publication NO. US20040067535A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kim, Jaeseob
 ; APPLICANT: Galant, Ron
 ; TITLE OF INVENTION: Alzheimer's Disease Linked Genes
 ; FILE REFERENCE: LSD-07417
 ; CURRENT APPLICATION NUMBER: US/10/263,929
 ; CURRENT FILING DATE: 2002-10-03
 ; NUMBER OF SEQ ID NOS: 213
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 144
 ; LENGTH: 3551
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-10-263-929-144
 Alignment Scores:
 Pred. No.: 0.00216 Length: 3551
 Score: 176.00 Matches: 187
 Percent Similarity: 30.63% Conservative: 93
 Best Local Similarity: 20.46% Mismatches: 304
 Query Match: 3.75% Indels: 332
 DB: 12 Gaps: 38
 US-09-787-016A-1 (1-2610) x US-10-263-929-144 (1-3551)
 QY 248 TCAGTGTGAGCAGGTATGGACAAAGGCGACCCGACCAATGAGGAGGCACCTAAGG 307
 Db 1761 SerValValGluLysProValThrArgLysSerGlyArgLysLeuLysLeuLys 1780
 QY 308 CCATCAACACCCACCA----- 322
 Db 1781 ArgSerSerProArgGlyGluAlaGlnLysLeuLeuLysMetGluAlaGlu 1800
 QY 323 -----GCAAGACTTCAGG-----AAACATGGGTTTTCGAA 355
 Db 1801 LysIleThrArgThrAlaSerLysSerSerGlyAspThrGluHisProGluProSer 1820
 QY 356 GGACCACTATGCGCAAGCGAGAGGGGCGCAGG-----GAC 390
 Db 1821 LeuProLeuSerArgSerArgArgAsnValArgSerValTyrrAlaThrMetThrAsp 1840
 QY 391 GCGGAGGCT-----GACCCACTGAGCCGCCA----- 417
 Db 1841 HisGluSerArgSerProAlaLysGluProValGlnProArgValThrArgLysArg 1860
 QY 418 -----CCCCACAGCAGCAGCAGCTGGGCGCTGTCCTG 447
 Db 1861 LeuGluArgGluLeuGlnGluAlaValProPro-----ThrThr 1874
 QY 448 CGCGCAGTGGGAGGAGCCCAAGCGCAGCTGAGCGGTGGAGCAGTTCCTGACCATTGGG 507
 Db 1875 ProArgArgGlyArgProProLysThrArgArgAlaGlu----- 1889
 QY 508 CGCGCGCGCGCAGGAGGACCATGCTCTCCCTGAGGATTCTGCTGAGCCCACTCC 567
 Db 1890 AspGlyGluHisGluArgLysGluProAlaGluThrProArgProAlaGluGlyTrpArg 1909

QY 568 TGCCCGCCACAGACGCGCAGACAGCAGCTCCGAGGCG----- 603
 Db 1910 SerProArgSerGlnLysSerAlaAlaAlaAlaGlyProGlnGlyLysAsGlyArgAsn 1929
 QY 604 -----AGCTGGAAACGGCTTCTGAG-----ACGAGAGCGGC--- 636
 Db 1930 GluGlnLysValGluAlaAlaAlaGluAlaGlyAlaGlnAlaSerThrArgGluGlyAsn 1949
 QY 637 CCCAGTCTGCTCCACAGCTGTGAAGGAACGACGACCTCTTCTGAAAAGGTGAAGGA 696
 Db 1950 ProLysSerArgGly-----GluArgGluAlaAlaSerGluProLysArg--- 1964
 QY 697 GGGGATGACCAACGATGACACCTCCGAT---AGTCACAGCGATGCCCTGACCTTGAAGAG 753
 Db 1965 ---AspArgArgAspProSerThrAspLysSerGlyProAspThrPheProValGluVal 1983
 QY 754 CTTTCAGATCGCTTCGAGGAACCGGNAACAGAGCCCACTGAGAGCCCTCGAAAGGG 813
 Db 1984 Leu-----GluArgLysProGluLysThrTyrrLysSer 1995
 QY 814 ATCCAGAGTCCCTCGGGAAGACGCG----- 840
 Db 1996 LysArgGlyArgAlaArgSerThrArgSerAlaMetAspArgAlaAlaHisGlnArgSer 2015
 QY 841 -----CGGAGGAGGCTCCCGCCGAG 861
 Db 2016 LeuGluMetAlaAlaArgAlaAlaGlyGlnAlaAlaAspLysGluAlaGlyProAlaAla 2035
 QY 862 ACTGTGGCTCCGAGGCGCAGTGTGAGGGCGTCTGCGCCAGTAAGCAG----- 915
 Db 2036 AlaSerProGlnGluSerGluSerProGlnLysGlySerGlySerSerProGlnLeuAla 2055
 QY 916 -----GAGCCCGACACGATCAGGGGTGTCTCCAGGCT 951
 Db 2056 AsnAsnProAlaAspProAspArgGluAlaGluGluSerAlaSerAlaSerThrAla 2075
 QY 952 GCGAAAGATGAC-----AGACAGAGTAGTTGGAGGAAAGCGCGCTCAGAC 999
 Db 2076 ProProGluGlyThrGlnLeuAlaArgGlnIleGluLeuGlu---GlnAlaValGlnAsn 2094
 QY 1000 ATC---AAAGATGAGGCGCTGGAGACTTGGGC----- 1029
 Db 2095 IleAlaLysLeuProGluProSerAlaAlaAlaAlaSerLysGlyThrAlaThrAlaThr 2114
 QY 1030 -----CGACCGAAGCTGAATGTGAGGTACGACCCCAACCGCTGTATTGC 1077
 Db 2115 AlaAlaSerGluGluProAlaProGlu---HisGlyHisLysPro----- 2128
 QY 1078 ATTTCCCGCCAGCTCAACACACAGGTTTANGATTGCTGTGACCCGCTGTGAAGAATGG 1137
 Db 2129 -----AlaHisGlnAlaSerGluThr 2135
 QY 1138 TTTTCATGCGGATTGTGTGGC-----ATTTCGAGGCTCGAGGGAGGCTTTTGGAAAGG 1191
 Db 2136 GluLeuAlaAlaAlaIleGlySerIleIleSerAspAlaSerGly----- 2150
 QY 1192 AATGGGAGAGCTATATCTGCCCA----- 1215
 Db 2151 GluProGluAsnPheSerAlaProProSerValProProGlySerGlnThrHisProArg 2170
 QY 1216 -----AACTGCACCATTCGCAAGTCAGGATGAG 1245
 Db 2171 GluGlyMetGluProGlyLeuHisGluAlaGluSerGlyIleLeuGluThrGlyThrAla 2190
 QY 1246 ACTCATTCAGAA-----ACGCGACATCAG 1269
 Db 2191 ThrGluSerSerAlaProGlnValSerAlaLeuAspProProGluGlySerAlaAspThr 2210
 QY 1270 CAGGAAGCTAATGAGACCTGAGATCTGATGCCCGGATTGTACAAGTATAGGAACA 1329
 Db 2211 LysGluThrArgGlyAsnSerGlyAspSer-----Val 2221
 QY 1330 ATGAGACGACAGTCTAGC-----GAAGACCGAGGGATAAG 1365


```
DB 10118 ArgGlyAlaGlyThr-----SerArgSerSerArgCysArgSer 10130
QY 784 GTTCCCGCTTCTCGAAGCGGATCTGAAGCTCTTTCAAGGTCAGGCCATCGCTGCAC 725
DB 10131 ThrProAsnSerCysAlaAlaThrSerPro-----ArgSerArg 10143
QY 724 TATCGGAGGCTCATCGTGTCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 686
DB 10144 TyrglyArgProProAlaGProProProProProProProProProProProPro 10163
QY 685 TTTCAGAAG-----AGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 647
DB 10164 ThrProAlaThrAlaAlaCysAlaAlaValArgThrThrAlaProThrArgProTrpLys 10183
QY 646 CAGACTGGGGCCCG-----TTCTGCTCTCAGAGCGCTTTTCCAGCTGCTCTCGG--- 596
DB 10184 ArgSerArgTrpThrProTyrrValTrpThrThrSerSer-----ProArgThrProAlaTrp 10202
QY 595 AGGCTGTCTCGGCTCTCTGCGGCGGAGGAGCGCTGCTCTCTCTCTCTCTCTCTCTCT 539
DB 10203 ThrTrpSerArgTrpThrTrpArgAlaAlaArgTrpAla-----ProCysAlaGly 10219
QY 538 AGACAGGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 485
DB 10220 ArgProSerCys-----CysAlaGlySerProProTrpTrpSerSerAsnThrAlaAla 10237
QY 484 CGCGCTCAGTGCCTCT-----TGG----- 467
DB 10238 ThrThrArgCysAlaThrThrAlaProProAlaThrThrCysGlyArgCysTrpSerThr 10257
QY 466 ---GCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 440
DB 10258 ThrSerAlaThrArgCysThrProCysArgAlaGlyTrpArgArgSerArgAlaSerThr 10277
QY 439 GGGCCCA-----GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 395
DB 10278 GlyProLeuSerProProSerCysAlaArgSerGlyThrSerSerProAlaAlaGlyPro 10297
QY 394 CCGGCTCCCTGCGCCCT 377
DB 10298 CysProProValProPro 10303

RESULT 9
; Sequence 24 Application US/09823187
; Publication No. US20030096952A1
; GENERAL INFORMATION:
; APPLICANT: Burgess, Catherine
; APPLICANT: Gusev, Vladimir Y
; APPLICANT: Liu, Xiaohong
; APPLICANT: Majumder, Kumud
; APPLICANT: Padigaru, Muralidhar
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Spaderna, Steven K
; APPLICANT: Spytek, Kimberly
; APPLICANT: Taupier, Raymond J
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-745
; CURRENT APPLICATION NUMBER: US/09/823,187
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/193,339
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/193,205
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/195,343
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: 60/195,088
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,005
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/195,792
```

```
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: 60/196,556
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: 60/197,081
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/197,525
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/197,087
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 103
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-823-187-24

Alignment Scores:
Pred. No.: 0.00266 Length: 676
Score: 172.00 Matches: 117
Percent Similarity: 34.32% Conservative: 80
Best Local Similarity: 20.38% Mismatches: 215
Query Match: 3.67% Indels: 162
DB: 10 Gaps: 23

US-09-787-016A-1 (1-2610) x US-09-823-187-24 (1-676)
QY 235 AGTGTCCAGCTTTTCAGTGTGGAGCAGGTATGACGACCAAGCGGACCCGAGCAATGAG 294
DB 130 AlaValThrAlaThrAlaAlaSerAspArgMetGluSerAspSerAspLysSer 149
QY 295 GAGGACCTTAAGCCATCAACCCACCAGCAAGAGTTCAGGAAACATGGGTTTCGA 354
DB 150 SerAspAsnSerGlyLeuLys-----ArgLysThrProAlaLeu--- 162
QY 355 AGGACCACTATCGCAAGCGAGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 414
DB 163 LysValSerValSerLysArgAlaArgLysAlaSerSerAspLeuAspGlnAlaSerVal 182
QY 415 CCACCCCAACAGCAGCTGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 474
DB 183 SerProSerGluGluGluAsnSerSerGluSerSerSerGluSerGluLysThrSerAspGln 202
QY 475 ACTGAGCGCGTGAGCAGTCTCTGACCATTT---GCGGCGCGCGCGCGG--- 519
DB 203 AspPheThrProGluLysLysAlaAlaValArgAlaProArgArgGlyProLeuGlyGly 222
QY 520 ---AGGAGGAGCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 570
DB 223 ArgLysLysLysLysAlaProSerAlaSerAspSerAspSerLysAlaAspSerAspGly 242
QY 571 ---CCGCGCACAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 618
DB 243 AlaLysProGluProValAlaMetAlaArgSerAlaSerSerSerSerSerSerSerSer 262
QY 619 TCT-----GAGACCAAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 663
DB 263 SerSerAspSerAspValSerValLysProProArgGlyArgLysProAlaGluLys 282
QY 664 ---GAACCAACAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 687
DB 283 ProLeuProLysProArgGlyArgLysProLysProLysProLysProLysProLysPro 300
QY 688 GTGAAAGAGGAGGAGTACAGCATGACACCTCCAGTAGTCAGCAGCGGAGCGGAGCGGAG 744
DB 301 ---SerSerSerAspSerAspSerAspSerAspSerAspSerAspSerAspSerAspSer 312
QY 745 TTGAAAGAGCTTCAGATCGC-----CTTCCAGGAAG 777
DB 313 IleSerGluTrpLysArgArgArgArgGluAlaArgGluAlaArgArgArg 332
QY 778 CGGAAACAGGAGCCCACTGAGAGCGCCCTGAAAGGAGTCCAGAGTCCCTCGCGGAGAGAG 837
```


Db 3 ArgSerGluPheProGly-----ArgProThrArgProAlaValThrAla 17
 QY 247 TTCAGTGTGGCAGAGTATGACGACAAAGCGCACCGAGCAATGAGGCGCACTAAG 306
 Db 18 ThrAlaAlaSerAspArgMetGluSerAspSerAspSerAspSerAspSer 37
 QY 307 GCCATCAAAACCCAGCAGCAAGAGTTCAGGAAACATGGGGTTTCGAGAGCACCATC 366
 Db 38 GlyLeuLys-----ArgLysThrProAlaLeu---LysMetSerVal 50
 QY 367 GCCAAGCAGAGGGCGCAGGAGCCGCGAGGCTGACCCACTCGAGCCGCCACCCACAG 426
 Db 51 SerLysArgAlaArgLysAlaSerSerAspLeuAspGlnAlaSerValSerProSerGlu 70
 QY 427 CAGCAGCTGGGCTCTCCCTGGCGCGAGTGGAGGCGAGCCCAACCGCAGCTGAGCGGTG 486
 Db 71 GluGluAsnSerGluSerSerGluSerGluLysThrSerAspGlnAspPheThrPro 90
 QY 487 GAGCAGTTCCTGACCATTT---CGCGGGCGCCCGCGC-----AGGAGGAGC 528
 Db 91 GluLysLysAlaAlaValArgAlaProArgArgGlyProLeuGlyGlyArgLysLysLys 110
 QY 529 ANGCGCTGCTCCCTCGAGGATTCTGGTGAGCCCGACGTCCTGCG----- 570
 Db 111 LysAlaProSerAlaSerAspSerAspSerLysAlaAspSerAspGlyAlaLysProGlu 130
 QY 571 CCGCCACAGACCGCAGACCTCCGAGGCGCAGCGTGGAAAGCGCTTCT 621
 Db 131 ProValAlaMetAlaArgSerAlaSerSerSerSerSerSerSerSerSerSer 150
 QY 622 -----GAGACAGAGCGCGCCCGAGTCTGCTCCACAGCTGTGAAG----- 663
 Db 151 AspValSerValLysLysProProArgGlyArgLysProAlaGluLysProLeuProLys 170
 QY 664 -----GAACACAGCAGCTCTTCTGAAAAGGTGAAAGAGAGG 699
 Db 171 ProArgGlyArgLysProLysProGluArgProProSerSer----- 184
 QY 700 GATGACCAAGTACACCTCCGATGATGACAGGATGGCGCTG---ACCTTGAAGAGCTT 756
 Db 185 -----SerSerSerAspSerAspSerAspLysValAspArgLysSerGluThrP 200
 QY 757 CAGAAATCGC-----CTTCGAGGAGCGCGGAACAGGAG 789
 Db 201 LysArgArgAspGluAlaArgArgArgGluLeuAlaArgArgArgGluGlu 220
 QY 790 CCACTGAGAGGCCCTGTAAGAGGATCCAGAGTCGCTCCGGAAGAGCGCCGAG--- 846
 Db 221 GluGluLeuArgArgLeuArgGluGluGluLysGluLysGluArgArgGluArg 240
 QY 847 -----GAGGTCGCCCGAG-----ACTGTGGCTCCGAGCCAGTGCACACTGTGGAG 894
 Db 241 AlaAspArgGlyGluAlaGluArgGlySerGlySerGlyAspGluLeuArgGlu 260
 QY 895 GCGCTCTCCCTAGTAAGCAGGAGCCCGAAGACGATCAGGGGGTTGTCCCGAGGCTGGG 954
 Db 261 AspAspGluProValLysLysArgGlyArgLysGlyArgGlyProProSerSer 280
 QY 955 AAGAT---GACAGAGAGTAACTGGGGGAAAGCGGCTCAGGACATCAAGATGAG 1011
 Db 281 SerAspSerGluProGluAlaGluLeuArgGluAlaLysSerLysLysLys 298
 QY 1012 GAGCCTGGAGACTTGGCGCGACCGAGACCTGAAATGTGAGGGTTACAGCCCGACCGCTG 1071
 Db 299 -----LysProGlnSerSerSerThrGluPro----- 307
 QY 1072 TATGCAATTTGCGCAGGCTCAACACAGAGTTTATGATTTGCTGTGACCGCTGTGAA 1131
 Db 308 -----AlaArgLysProGly----- 312
 QY 1132 GAATGTTTTCAGCGATTGTGGGGCAATTCTCAGGCTCGAGGAGGCTTTTGAAGAG 1191
 Db 312 ----- 312

QY 1192 AATGGGAGAGACTATATCTCCCAAACTGCACCAATTCGCAAGTGCAGGATGAGACTCAT 1251
 Db 313 -----GlnLysGluLysArg 317
 QY 1252 TCAGAAAGCGCAGATCAGCAGAGCTAAATGGAGACCTCGAGATGCTGATGGCACCGAT 1311
 Db 318 ValArgProGluGluLysGlnAlaLys-----ProValLysValGluArgThr--- 334
 QY 1312 TGTACAATATAGGAACATAGACAGAGTCTTAGCGAAGCAAGGATTAAGGGTGA 1371
 Db 335 -----ArgLysArgSerGluGlyPheSerMetAspArgLys 346
 QY 1372 ATTGAGAAGCTGCAATCCAAAGTGGCAAGAGAACTCAAG----- 1413
 Db 347 ValGluLysLysLysGluProSerValGluGlnLysLeuGlnLysLeuHisSerGluLe 366
 QY 1414 ATCTTCCAGCCTGTGATAGAGCGCCTGGTGCCTCAAAATGCTATTTGGCCCGGTGTGT 1473
 Db 367 LysPheAlaLeuLysValAspSerProAspValLysArgCysLeuAsnAla----- 383
 QY 1474 CAGTGGCGGACCGGACTCGTGTACTGCAAGTAACTGATCTATCTCTCAACAC----- 1527
 Db 384 ---LeuGluGluLeuGlyThrLeuGlnValThrSerGlnLeuLeuGlnLysAsnThrAsp 402
 QY 1528 GCGCAGCGCAATGAAGTTTCTA-----AGCTCAGGTAAAGACAGAGCCAAAG 1578
 Db 403 ValValAlaThrLeuLysLysLysLeuArgArgTyrlLysAlaLeuLysAspValMetGluLys 422
 QY 1579 CCTAAGAAAAGATGAGATGAGCAGAGAGCCAGAGCCAGCTCTCCGAATGGGTCTCAG 1638
 Db 423 AlaAlaGluValTyThrArgLeuLysSerArgValLeuGlyProLysLysLeuAlaVal 442
 QY 1639 CGAGTATTAATAATCTCTTCTGTGCACAGAGACAGCTCCAGAAAAAAGAGAGACCACA 1698
 Db 443 GlnLysValAsnLysAlaGlyMetGluLysGluLysAlaGluGluLysLeu----- 459
 QY 1639 GTGAAGAGGAGCGTGTGTCTCCGCGGAGTGAAGCACTCGGGAAGAGAGCAGCTTGT 1758
 Db 460 -----AlaGlyGluGluLeuAlaGlyGluGluAlaProGln 471
 QY 1759 GAGAGCAGCAGCGCTGCTGGCGGAGCGAT 1788
 Db 472 GluLysAlaGluAspLysProSerThrAsp 481

RESULT 12

US-10-203-708-28
 ; Sequence 28, Application US/10203708
 ; Publication No. US20030149238A1
 ; GENERAL INFORMATION:
 ; APPLICANT: SMITHKLINE BEECHAM CORPORATION
 ; APPLICANT: SMITHKLINE BEECHAM P.L.C.
 ; TITLE OF INVENTION: NOVEL COMPOUNDS
 ; FILE REFERENCE: GP50013
 ; CURRENT APPLICATION NUMBER: US/10/203,708
 ; CURRENT FILING DATE: 2002-08-13
 ; PRIOR APPLICATION NUMBER: PCT/US01/04703
 ; PRIOR FILING DATE: 2001-02-14
 ; PRIOR APPLICATION NUMBER: 60/182,172
 ; PRIOR FILING DATE: 2000-02-14
 ; PRIOR APPLICATION NUMBER: 60/186,084
 ; PRIOR FILING DATE: 2000-02-29
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 28
 ; LENGTH: 676
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-203-708-28

Alignment Scores:
 Pred. No.: 0.00313 Length: 676
 Score: 171.00 Matches: 117

; PRIOR APPLICATION NUMBER: 60/195,343
 ; PRIOR FILING DATE: 2000-04-05
 ; PRIOR APPLICATION NUMBER: 60/195,088
 ; PRIOR FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: 60/195,005
 ; PRIOR FILING DATE: 2000-04-06
 ; PRIOR APPLICATION NUMBER: 60/195,792
 ; PRIOR FILING DATE: 2000-04-10
 ; PRIOR APPLICATION NUMBER: 60/196,556
 ; PRIOR FILING DATE: 2000-04-11
 ; PRIOR APPLICATION NUMBER: 60/197,081
 ; PRIOR FILING DATE: 2000-04-13
 ; PRIOR APPLICATION NUMBER: 60/197,525
 ; PRIOR FILING DATE: 2000-04-14
 ; PRIOR APPLICATION NUMBER: 60/197,087
 ; PRIOR FILING DATE: 2000-04-14
 ; NUMBER OF SEQ ID NOS: 103
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 87
 ; LENGTH: 678
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-09-823-187-87

Alignment Scores:
 Pred. No.: 0.00313 Length: 678
 Score: 171.00 Matches: 135
 Percent Similarity: 32.25% Conservatives: 83
 Best Local Similarity: 19.97% Mismatches: 194
 Query Match: 3.65% Indels: 264
 DB: 10 Gaps: 32

US-09-787-016a-1 (1-2610) x US-09-823-187-87 (1-678)

QY 214 GRATTAGGATTTCAGGAAAGTGTCCAGCTTTCAGTGTGGAGCAGGTATGAGCAGC 273
 Db 125 ValMetThrValThrAlaValThrThrAlaThrSerAspArgMetGluSerAspSer 144
 QY 274 AAAGGCGACCCGACCAATGAGGAGGACCTTAAGGCCATCAAAACCACCAAGAGTTC 333
 Db 145 AspSerAspLysSerSerAspHisSer----- 153
 QY 334 AGGAAACATGGGGTTTTCGAGGACCACT-----ATGCCCAAGCGA 375
 Db 154 -----GlyLeuLysArgLysThrProValLeuLysValSerValSerLysArg 169
 QY 376 GAGGCGCAGGAGACCGGAGGCTGACCCACTGGAGCCGCCACCCACAGCAGCAGCTG 435
 Db 170 AlaArgAlaSerSerAspLeuAspGlnAlaSerValSerProSerGluGlu----- 187
 QY 436 GGCCTGTCCCTGCGCGCAGTGGGAGGCGAGCCCAAGCGCACTGAGCGCGTGGAG----- 489
 Db 188 -----AspSerGluSerProSerGluSerGluSerLysThrSerAspGln 201
 QY 490 -----CAGTTCTGACCATTCGCGG-----CGCGCGGAGGAGGAGCATG 531
 Db 202 AspPheThrProGluLysLysThrAlaAlaArgProArgArgGly----- 217
 QY 532 CCGTGTCTCCCTGAGGATTCTGTGTGAGCCACG-----TCCGCGCCCGCC----- 576
 Db 218 ProLeuGlyArgLysLysLysHisProThrGlyTyraLysProGlnLysValPro 237
 QY 577 -----ACAGACCGCGAGACAGCCTCCGAGCGAGCGGTGAA----- 612
 Db 238 SerAlaSerAspSerLysAlaAspSerAspGlyAlaLysGluGluProValVal 257
 QY 613 -----AGCGCTTCTGAGACGAGCGGCCCGCCCGCTGCT 648
 Db 258 ThrAlaGlnProSerProSerSerSerSerSerSerSerSerSerSerSerSer 277
 QY 649 TCCACAGCTGTGAAG----- 663
 Db 278 AspValSerValLysLysProProArgGlyArgLysProAlaGluLysProProLys 297

QY 664 -----GAACGACACAGCCTCTCTGAAAAGGTGAAAGAGGG 699
 Db 298 ProArgGlyArgProLysProGluArgProProSerThr----- 311
 QY 700 GATGACACGATGACACCTCGATAGTACAGCGCATGCGCTGACCTGAAAGAGCTTCAG 759
 Db 312 -----SerSerSerAspSerAspSerGlyGluValAspArgLys 327
 QY 760 AAT-----CGCTTCGAGGAGCGGAA 783
 Db 328 GluTrpLysArgArgAspGluLysArgGluLysGluAlaArgArgArgGlu 347
 QY 784 CAGAGCCCACTGAGAGCCCGCTGAAAGGATCAGAGTCGCTCGCGAAGAGCGCGG 843
 Db 348 GlnGluGluGluLeuArgLysGlu-----GlnGluArgGluGluLysGluArg 366
 QY 844 GAGAGGGTCCCGCGAGACTGTGGCTCCGAGGCCAGTGCACCTGTGGAGGCGCTCTG 903
 Db 367 LysGlu-----ArgAlaGluArgGlySer----- 375
 QY 904 CCCAGTAGCAGGAGCCCGAGAACGATCAGGGGTGTGTCCAGGCTGGGAAAGATGAC 963
 Db 376 ---SerGlyGluGluLeuGluAspGluProVal-----LysLys 388
 QY 964 AGAGAGAGTAAAGTTGGAGCGGAAAGCGCTCAGGACATCAAGATGAGAGCCT---GGA 1020
 Db 389 ArgSerArgLysAlaArgGlyArgGlyThrProSerSerSerAspSerGluProGluGly 408
 QY 1021 GACTTGGGCCGA-----CCGAAAGCTGAATGTGAGGGTTACAGCCCC 1062
 Db 409 GluLeuGlyLysGluGlyLysLysLysLysLysSerGlnLeuProGlySerGluSer 428
 QY 1063 AACGCCCTGATTGCTTTGCGCGCAGCTCACAACAACAGGTTTATGATTGCTGTGAC 1122
 Db 429 -----AlaArgLysPro----- 432
 QY 1123 CGCTGTGAAGAATGTTTTCATGGCGATTGTGTGGCATTTCTGAGGCTCGAGGAGGCTT 1182
 Db 433 -----GlyGlnLysGluLysArgGlyArg--- 440
 QY 1183 TTGGAAGGAATGGGAAGACTATATCTGCCAAACTGCACCATCTTCAAGTGCAGGAT 1242
 Db 440 ----- 440
 QY 1243 GAGACTCATTAGAAACCGCAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCTAT 1302
 Db 441 -----ProAspGluLysProArgAlaArgProValLysValGlu 453
 QY 1303 GGCACCGATTGTACAGTATAGGAACAATAGCAGAGAAGTCTAGCGAGACCAAGGATA 1362
 Db 454 ArgThr-----ArgLysArgSerGluGlyLeuSerLeu 464
 QY 1363 AAGGCTAGAAATGAGAAAGCTCAAAATCCAACTGGCAAGAGAACTCAAG----- 1413
 Db 465 GluArgLysGlyGluLysLysLysGluProSerValGluArgLysGlnLysLeuHis 484
 QY 1414 -----ATCTCCAGCCTGTGTAGAGCGCGCTGTGCTCAAAATGTATGTGCCCC 1464
 Db 485 SerGluIleLysPheAlaLeuLysValAspAsnProAspValArgLysCysLeuSerAla 504
 QY 1465 -----GGTGTGTGTACCTGGCG-----CAGCCCGACTCG 1494
 Db 505 LeuGluGluLeuGlyThrLeuGlnValThrSerGlnIleLeuGlnLysAsnThrAspVal 524
 QY 1495 GTG-----TACTCAGTAATGACTGTATCTCAACACAGCC 1530
 Db 525 ValAlaThrLeuLysLysIleArgArgTyrLysAlaAsnLysAspValMetAlaLysAla 544
 QY 1531 GCA-----CCGCAATG 1542
 Db 545 AlaGluValTyrThrArgLeuLysSerArgValLeuGlyProLysValGluAlaLeuGln 564

```
QY 1543 AAGTTCTAAGCTCAGTAAAGAACACAGAGCCAAAGCCTAAAGAAAGATGAAGATGAG 1602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 LysValAsnLysAlaGlyAlaGluLysGluArgAlaAspAsnGluLysLeuGluGln 584
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1603 CAGAGAACGCCAGTCTTCGAAATCGCTGCTCAGCAGGATTAATAATCTCTCTCTG 1662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 585 Pro-----GlyGluAlaProArgGluLeuAlaGluAsp 596
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1663 CAC-----AAGACACAGCTCCAGAAAGAGAGACACACAGTGAAGAGGCGAGT 1713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 597 GluProSerThrAspArgSerAlaProValAsnGlyGluAlaThrSerGlnLysGlyGlu 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1714 GTGTCCTCGCGGAGTGAAGACCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 AspMetGluAspArgAlaGlnGluAspGlyGlnAspSer-----GluAspGlyProArg 634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1774 TCGTGGCGAGCAGCATCAC-----AATTACAAATGACAGTAAAGCCA 1812
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 GlyGlySerSerGluLeuHisAspSerProArgAspAsnSerAspProAlaLysPro 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1813 -----GAAAGACTGCTGCTCCCTCGCCGTC 1839
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 GlyAsnGluArgGlnAspHisGluArgThrArgLeuAlaSerGluSer 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-09-863-776-52
; Sequence 52, Application US/09863776
; Publication No. US20030198953A1
; GENERAL INFORMATION:
; APPLICANT: Spytex, Kimberly A
; APPLICANT: Majumder, Kumud
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Mishra, Vishnu
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderna, Steven K
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Rastelli, Luca
; APPLICANT: Li, Li
; APPLICANT: Taupier, Raymond J
; APPLICANT: Gangoli, Esha
; TITLE OF INVENTION: No. US20030198953A1el Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21404-020
; CURRENT APPLICATION NUMBER: US/09/863, 776
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 09/540, 763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/206, 679
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206, 688
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/206, 829
; PRIOR FILING DATE: 2000-05-24
; PRIOR APPLICATION NUMBER: 60/207, 748
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/207, 798
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/208, 263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/208, 831
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/209, 451
; PRIOR FILING DATE: 2000-06-05
; PRIOR APPLICATION NUMBER: 60/210, 060
; PRIOR FILING DATE: 2000-06-07
; PRIOR APPLICATION NUMBER: 60/219, 507
; PRIOR FILING DATE: 2000-07-20
; PRIOR APPLICATION NUMBER: 60/221, 337
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/221, 927
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/263, 135
; PRIOR FILING DATE: 2001-01-19
; PRIOR APPLICATION NUMBER: 60/263, 688
```

```
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/263, 694
; PRIOR FILING DATE: 2001-01-24
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 678
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-863-776-52

Alignment Scores:
Score: 0.00313 Length: 678
Pred. No.: 171.00 Matches: 135
Percent Similarity: 32.25% Conservative: 83
Best Local Similarity: 19.97% Mismatches: 194
Query Match: 3.65% Indels: 264
DB: 10 Gaps: 32

US-09-787-016A-1 (1-2610) x US-09-863-776-52 (1-678)
QY 214 GTATTAGATTTCAGGGAAGAGTCCAGCTTTCAGTGTGGAGCAGGTATGGACGAC 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 ValMetThrValThrAlaValThrThrAlaThrSerAspArgMetGluSerAspSer 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 274 AAAGCGGACCCGACCAATGAGGAGGACGACCTAAGCCCATCAAAACCCAGCAAGAGTTTC 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 145 AspSerAspLysSerSerAspHisSer----- 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 334 AGGAAACATGGGGTTCGAGAGGACCACT-----ATGCCCAAGCGA 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 -----GlyLeuLysArgLysThrProValLeuLysValSerValSerLysArg 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 376 GAGGCGGAGGCGGAGGAGGCTGACCCACTGAGAGCGGACCCACCCAGCAGCAGCAGCTG 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 AlaArgArgAlaSerSerAspLeuAspGlnAlaSerValSerProSerGluGlu----- 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 436 GGCCTGTCCCTGCGCGGAGTGGGAGGCGGACCCAGCCACCTGAGCGGCGTGCAG----- 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 -----AspSerGluSerProSerGluSerGluLysThrSerAspGln 201
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 490 -----CACTTCTCTGACCATTCGCGG-----CGCGCGGCGGAGGAGGAGCAGT 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 202 AspPheThrProGluLysLysThrAlaAlaArgProProArgArgGly----- 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 532 CCTGTCTCCCTCGAGGATTCGTGAGCCGACG-----TCTGCCCCGCC----- 576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 218 ProLeuGlyGlyArgLysLysLysHisProThrGlyfYrAlaCysProGlnLysValPro 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 577 -----ACAGAGCCGAGACAGCTCCGAGGCGGAGCGTGGAA----- 612
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 SerAlaSerAspSerAspSerLysAlaAspSerAspGlyAlaLysGluProValVal 257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 613 -----AGCGCTTCTGAGACCAAGAGCGGCCCCCAGTCTGTCT 648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 258 ThrAlaGlnProSerProSerSerSerSerSerSerSerSerSerSerSerSerSer 277
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 649 TCACAGCTGTGAG----- 663
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 278 AspValSerValLysLysProProArgGlyArgLysProAlaGluLysProProLys 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 664 -----GAAACGACGAGCTCTTCTGAAAGGTGAAGAGAGCG 699
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 ProArgGlyArgArgProLysProLysProLysProLysProLysProLysProLys 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 700 GATGACACGATGACACTCCGATAGTACACGATGGCTGACCTTGAAGAGCTTTCAG 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 -----SerSerSerAspSerAspSerGlyGluValAspArgLysSer 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 760 AAT-----CGCCTTCGAGGAGCGGGA 783
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 GluTrpLysArgArgAspGluArgArgGluLeuGluAlaArgArgArgGlu 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


QY 619 TCT-----GAGACAGAGCGCCGCCAGTCTGTCTCCACAGCTGTGAAG 663
Db 304 SerSerAspSerAspValSerValLysLysProProArgGlyArgLysProAlaGluLys 323
QY 664 -----GACGACCGACCTCTTCTGAAAG 687
Db 324 ProLeuProLysProArgGlyArgLysProLysProGluArgProProSerSer 341
QY 688 GTCAAAGGAGGGGATGACACCGATGACCTCCGATAGTACAGCGATGCCCTG--ACC 744
Db 342 -----SerSerSerAspSerAspSerAspGluValAspArg 353
QY 745 TTCAAAGAGCTTCAGAAATCCG-----CTTCGCGGGAAG 777
Db 354 IleSerGluTrpLysArgArgAspGluAlaArgArgGluLeuGluAlaAlaArgArgArg 373
QY 778 CGGACACAGAGCCCACTGAGGCGCCCTGAAAGGATCCAGAGTCGCTCGGGAAGAG 837
Db 374 ArgGluGlnGluGluLeuArgArgLeuArgGluGlnGluLysGluGluLysGluArg 393
QY 838 CGCGCGGAG-----GAGGCTCCCGCCGAG-----ACTGTGGGCTCCGAGGCCAGT 882
Db 394 ArgArgGluArgAlaAspArgGlyGluAlaGluArgGlySerGlyGlySerGlyAsp 413
QY 883 GACACTGTGGAGGCGCTCTGCCAGTAAAGCAGGAGCCCGAGAACATCAGGGGTTGTG 942
Db 414 GluLeuArgGluAspAspGluProValLysLysArgGlyArgLysGlyArgGly 433
QY 943 TCCAGCGTGGGAAGAT---GACAGAGAGAGTAAGTTGAGGGAAGCGGCTCAGGAC 999
Db 434 ProProSerSerSerAspSerGluProGluAlaGluLeuGluArgGluAlaLysLysSer 453
QY 1000 ATCAAGATGAGGAGCGCTGGAGACTTGGCGCGGCGGACCGAACCTGAATGTGAGGGTACGAC 1059
Db 454 AlaLys-----LysProGlnSerSerSerThrGlu 463
QY 1060 CCCAAGCCCTGTATTGCAATTTCCCGCCAGCTCACACACAGGTTTATGATTGCTGT 1119
Db 464 Pro-----AlaArgLysProGly----- 469
QY 1120 GACCGCTGTGAAGATGGTTTCATGCGCATTTGTGGGCATTTCTGAGGCTGAGGGAGG 1179
Db 469 ----- 469
QY 1180 CTTTTGGAAGGANTGGGAGACTATATCTGCCCCAAATGTCACCATTTCTGCAAGTGACAG 1239
Db 470 -----Gln 470
QY 1240 GATGAGACTCATTGAGAAACGGGAGATCAGCAGGAAGCTAAATGGAGACCTGGAGATGCT 1299
Db 471 LysGluLysArgValArgProGluGluLysGlnAlaLys-----ProValLysVal 488
QY 1300 GATGGCCCGGATTTGACAGTATAGGACAAATAGACGACAGGTTCTAGCGAGACCAAGG 1359
Db 489 GluArgThr-----ArgLysArgSerGluGlyPheSer 499
QY 1360 ATAAAGGGTAGAATGAGAAAGCTGCAAAATCCAAGTGGCAAGAGAACTCAAG----- 1413
Db 500 MetAspArgLysValGluLysLysLysLysLysLysLysLysLysLysLysLysLys 519
QY 1414 -----ATCTTCCAGCCTGTGATAGAGCGCTCGTGCCTCAAAATGATTGGC 1461
Db 520 HisSerGluIleLysPheAlaLeuLysValAspSerProAspValLysArgCysLeuAsn 539
QY 1462 CCGGGTGCTGTCAGTGGCGGCGCCGCTCGGTGACTGTCAGTAAATGACTGTATCCTC 1521
Db 540 Ala-----LeuGluGluLeuGlyThrLeuGlnValThrSerGlnIleLeuGln 555
QY 1522 AAACAC-----GCCGAGGAGCAATCAAGTTCTTA-----AGCTCAGGTAAAGAA 1566
Db 556 LysAsnThrAspValValAlaThrLeuLysLysIleArgGlyrLysAlaAsnLysAsp 575

QY 1567 CAGAAGCCAAAGCCTAAAGAAAGATGAAGTGAAGCCAGAGCCAGTCTTCCGAAA 1626
Db 576 ValMetGluLysAlaAlaGluValThrArgLeuLysSerArgValLeuGlyProLys 595
QY 1627 TCGGTCCTCAGGCGAGGTATTAAAATCTTCTGTGCACAGAGACCCAGCTCCAGAAAAA 1686
Db 596 IleGluAlaValGlnLysValAsnLysAlaGlyMetGluLysGluLysAlaGluGluLys 615
QY 1687 AAAGAGACCAACAGTGAAGAGGCGAGTGTGTCTCCCTGGCGGAGTGAAGCACTCGGAAG 1746
Db 616 Leu-----AlaGlyGluGluLeuAlaGlyGlu 624
QY 1747 GAAGCAGCTTGTGAGAGCAGCAGCCGCTGCTGGCGGAGCGAT 1788
Db 625 GluLeuAlaGlyGluGluAlaProGlnGluLysAlaGluAsp 638

Search completed: April 28, 2004, 11:22:28
Job time : 242.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2004, 10:49:27 ; Search time 29.5 Seconds
(without alignments)
9135.172 Million cell updates/sec

Title: US-09-787-016A-1
Perfect score: 4689
Sequence: 1 ctcggggcgctcgccacc.....actcttaagatcatatctg 2610

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delcp 6.0 , Delcxt 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/USPTO_spool_p/US09787016/runat_28042004_094330_26973/app_query.fasta_1.2759
-DB=Issued Patents AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=0 -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09787016 -CGN 1 1 29 -runat_28042004_094330_26973 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	174.5	3.7	2289	3	US-09-051-019-2
2	171	3.6	700	4	US-09-252-991A-32447
3	169	3.6	1404	4	US-08-801-308-1
4	164.5	3.5	1075	4	US-09-252-991A-18387
5	164.5	3.5	1427	4	US-09-252-991A-27005
6	163.5	3.5	1780	1	US-08-769-309A-5
7	163.5	3.5	1780	3	US-08-994-570-5
8	162.5	3.5	1805	1	US-07-853-913-2
9	162	3.5	341	4	US-09-252-991A-22507
10	161	3.4	238	4	US-09-257-179-80
11	160	3.4	8991	4	US-08-714-741-32
12	158.5	3.4	779	4	US-10-164-595-56

13	156.5	3.3	778	1	US-08-439-818A-5	Sequence 5, Appli
14	156.5	3.3	778	2	US-08-751-965-5	Sequence 5, Appli
15	156.5	3.3	778	2	US-08-738-975-5	Sequence 5, Appli
16	156.5	3.3	778	2	US-08-728-626-5	Sequence 5, Appli
17	156.5	3.3	778	3	US-08-808-599A-5	Sequence 5, Appli
18	154.5	3.3	1079	4	US-09-489-039A-7502	Sequence 7502, Ap
19	154.5	3.3	1618	1	US-07-853-913-4	Sequence 4, Appli
20	153.5	3.3	1596	4	US-08-978-277A-4	Sequence 5, Appli
21	153	3.3	3969	3	US-08-061-376-5	Sequence 5, Appli
22	152.5	3.2	442	4	US-09-252-991A-23285	Sequence 23285, A
23	152.5	3.2	632	4	US-09-252-991A-23129	Sequence 23129, A
24	152.5	3.3	735	4	US-10-164-595-80	Sequence 80, Appli
25	152.5	3.3	784	4	US-10-164-595-79	Sequence 79, Appli
26	152.5	3.3	843	4	US-10-164-595-54	Sequence 54, Appli
27	152.5	3.3	904	4	US-08-976-594-615	Sequence 615, App
28	152.5	3.3	1187	1	US-08-320-559-28	Sequence 28, Appli
29	152.5	3.3	1187	3	US-08-545-860D-28	Sequence 28, Appli
30	152.5	3.3	1187	5	PCT-US94-04496-28	Sequence 26, Appli
31	152.5	3.3	1210	1	US-08-320-559-26	Sequence 26, Appli
32	152.5	3.3	1210	3	US-08-545-860D-26	Sequence 26, Appli
33	152.5	3.3	1210	5	PCT-US94-04496-26	Sequence 726, App
34	152	3.2	2468	4	US-09-976-594-726	Sequence 19431, A
35	151	3.2	871	4	US-09-252-991A-19431	Sequence 29689, A
36	150	3.1	226	4	US-09-252-991A-29689	Sequence 14, Appli
37	147.5	3.1	1051	4	US-09-428-711A-14	Sequence 2, Appli
38	147.5	3.1	1061	4	US-09-762-481B-2	Sequence 28242, A
39	147	3.1	504	4	US-09-252-991A-28242	Sequence 30843, A
40	146.5	3.1	663	4	US-09-252-991A-30843	Sequence 5, Appli
41	146.5	3.1	732	1	US-08-317-522A-5	Sequence 27980, A
42	146.5	3.1	1002	4	US-09-252-991A-27980	Sequence 26065, A
43	145	3.1	836	4	US-09-252-991A-26065	Sequence 31638, A
44	144	3.0	253	4	US-09-252-991A-31638	Sequence 32910, A
45	144	3.0	469	4	US-09-252-991A-32910	

ALIGNMENTS

RESULT 1
US-09-051-019-2
; Sequence 2, Application US/09051019
; Patent No. 6103229
; GENERAL INFORMATION:
; APPLICANT: KAHMANN, Regine and QUADBECK-SERGER, Claudia
; TITLE OF INVENTION: Regulatory gene from Ustilago maydis
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Keil & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
; COMPUTER: IBM AT-compatible, Pentium processor
; OPERATING SYSTEM: Windows 98
; SOFTWARE: WordPerfect version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/051,019
; FILING DATE: 31-MAR-1998
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-051-019-2

Alignment Scores: 6.58e-06 Length: 2289
Pred. No.: 174.50 Matches: 120
Score: 34.28% Conservative: 85
Percent Similarity: 34.28% Mismatches: 184
Best Local Similarity: 20.07%

Query Match:	3.72%	Indels:	209
DB:	3	Gaps:	28
US-09-787-016A-1 (1-2610) x US-09-051-019-2 (1-2289)			
QY	346	GGTTTTTGAAGGACCACTATCCCAAGCGAGAGAGGCGCAGGGACGCGAGGCTGACCCA	405
DB	1387	GlySerArgArgLeuThrAspAlaAspArg-----AspGluGluLeuAsp----	1401
QY	406	CTGGAGCCGCCACCCCCACAGCAGCAGCTGGGCGCTGCTCCTCGGCGCAGTGGGAGCGAG	465
DB	1401	-----	1401
QY	466	CCCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACCATTCGCGCGCGCGCGCAGGAGG	525
DB	1402	---LysLeuValGluArgValGluAspAlaThrAspProAlaAspGlnAsnLysPro	1420
QY	526	AGCATGCTGTCTCCCTGAGGAGTTCGTGTAGCCACCGTCTCTGCGCCGCCACAGACGGC	585
DB	1421	AsnAlaArgAsnCysIleCysArgSerSerMetProIleAlaIleProSerSerGly	1440
QY	586	GAGACAGCTCCGAGGCGCCTGGAA-----AGCGTCTCAGACCAAGACGGCG	636
DB	1441	AlaGluCysSerArgCysArgValGlnTyrHisLeuSerCysIleLysValArgSerSer	1460
QY	637	CCCCAGTCT-----	651
DB	1461	GluValSerArgAlaGluGlyGlyTyrValCysProPheCysProTyrGlySerAla	1480
QY	652	ACAGCTGTGAAGAACGACCGACTCTCT-----GAAAGGTGAAGGAGG	699
DB	1481	ProPheLeuMetArgLysAlaIleSerIleAlaAspLeuSerLysLeuValTyrAsp	1500
QY	700	GATGACCA-----GATGACACTCCGATAGTCAGCAGCGCATGCGCTTGAAGAG	753
DB	1501	GlnAspHisArgArgAspGlnPheLysPheLeuProLeuGluTyrAspAlaIleGluGlu	1520
QY	754	CTTCAGAACTGCTTCGAGGAAGCGGAACAGAGCCCTCAGAGGCCCTTGAAGGG	813
DB	1521	ValVal-----AlaLysAlaLysArgPheGluThrAlaAlaLysArgMetIleLysThr	1538
QY	814	ATCCAGAGTCGCTGCGGAGAGAGCGCGGAGAG-----	849
DB	1539	Leu---SerLeuMetArgArgAspGlnLysGlnValIleLeuAlaHisTyrLeuArgArg	1557
QY	850	-----GATCCCGCCGAG-----	861
DB	1558	SerIleGlyCysProValAspValLeuGlyProGluLysValAsnMetLeuAspLeuIle	1577
QY	862	-----ACTGTGGCTCCGAGGCCAGTGACACT-----	888
DB	1578	SerGluAsnLeuAlaLeuGlySerGlnGlnClyAspAlaProMetAlaProVal	1597
QY	889	-----GTGGAGCGCGTCTGCCAGTAAGCAGAG-----	918
DB	1598	GluArgIleLysAlaSerThrProAlaArgSerAspGluArgThrGluThrThrPro	1617
QY	919	-----CCCGAGAACCATCAGCGGGTGTGTCTCCAG	948
DB	1618	LeuProArgSerSerArgValProAlaProAlaAspArgAspSerGly-----SerPro	1635
QY	949	GCTGGGAAATGACAGAGAGTAACTTCSAG---CGAAAGCGCTCAGGACATCAAA	1005
DB	1636	AlaValArgAspAspArgLysArgLysArgLysArgLysArgLysArgLysValPhe	1655
QY	1006	GATGAGGAGCCTCGAGACTTGGGCGCAGCGAAGCCTGAATGTGAGGTTTACGACCCCAAC	1065
DB	1656	GlnGluGluIleGly-----IleGlyAlaTyrArg-----AspArgGln	1668
QY	1066	GCCTGTATGTCATTTCGCCCGCCCTCACAACAAGTTTATGATTTCGTGTGACCGC	1125
DB	1669	ProIleTyrCysLeuCysHisGluProGluSerGlyArg---MetIleAlaCysAspLys	1687

```

RESULT 2
US-09-252-991A-32447
? Sequence 32447, Application US/09252991A
? Patent No. 6551795
? GENERAL INFORMATION:
? APPLICANT: Marc J. Rubenfield et al.
? TITLE OF INVENTION: NUCLEIC ACID AND AMINO
? FILE REFERENCE: 107196.136
? CURRENT APPLICATION NUMBER: US/09/252,991A
? CURRENT FILING DATE: 1999-02-18
? PRIOR APPLICATION NUMBER: US 60/074,788
? PRIOR FILING DATE: 1998-02-18
? PRIOR APPLICATION NUMBER: US 60/094,190
? PRIOR FILING DATE: 1998-07-27
? NUMBER OF SEQ ID NOS: 33142
? SEQ ID NO 32447
? LENGTH: 700
? TYPE: PRT
? ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32447

```

[illegible]

RESULT 2
US-09-252-991A-32447
; Sequence 32447, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; SUBJECT: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; SUBJECTS FOR DIAGNOSTICS AND THERAPEUTICS

; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-32447

Alignment Scores:

Pred. No.:	7.54e-06	Length:	7000
Assignment Scores:		Matches:	150
Score:	171.00	Conservative:	71
Percent Similarity:	31.44%	Mismatches:	276
Best Local Similarity:	21.34%	Indels:	206
Query Match:	3.65%	Gaps:	25
DB:	4		

US-09-787-016A-1 (1-2610) x US-09-252-991A-32447 (1-700)

188	QY	GGTTTCGTCTGC	CAGGGTTTTTGGTGTATTTAGATTC	CAGGAAAGTGTCCAAGCTT	247
19	Db	GlyAspValCysGln	AlaLeuAlaGluGlnValGlyGln-GlnGlnAspArgAla	Hi	38
248	QY	TCAGTGTGGAGCAGGTATG	GACACAAAGCGCACCCGAGCAATAGAGGACGACCACTAAGG	307	
38	Db	sGlnCys-----	AlaGlyGlnGlnArgHisProGlyArgAlaGly	51	
308	QY	CCATCAAAACCCAC	CAGCAAGAGTTCAGGAAACATGGGGTTTTTCGAAGGACCACTATCG	367	
51	Db	-----	-----	51	
368	QY	CCAAGCGAGAGGG-----	CGCAGGGGACCGGAGGCTGACCCACTCGAGCGCCACCC	421	
52	Db	-GlnAlaProGlyGlnLeu	ArgAlaGlyGlnGlyAspGluGlyAlaGlyArgG	71	
422	QY	CACAGCAGCAGCTGGGGCTCT	CCCTCGCGCGCAGTGGGACGACGCCCAAGCGCACTGAGC	481	
71	Db	yGlyGlyAspGlyGlyHis	AlaAspGlyAspGln-----	82	
482	QY	GCGTGGAGCAGTTCCTC	ACCATTGCGGGCGCGCGGACGAGGAGCATGCTGTCTCCC	541	
83	Db	-----	AspHisLeuGlnAlaHisAlaLeuHisArgHisAlaGluLeuG	97	
542	QY	TGGAGGA-----	TTCTGTGAGCCACAGCTCTGCCCGCCACAG	580	
97	Db	yGlyAlaGlyIleAlaGln	LeuGlnPheGlnAlaThrValValGlnArgGlnAr	117	
581	QY	ACGCCGAGACAGCCTCCG	AGCGGACGCTTTCGACACAGAAAGCGCTTTCGACACAGAAAGCGGCC	640	
117	Db	gHis-----	ValGlyGluGlnAlaGlnArgCysArgAspValArgPro--	132	
641	QY	AGTCTGCTTCCACAGCTGT	GAGAAACGACAGCAGCCTCTTCTGAAAAAGGT-----GA	691	
133	Db	-----	-----GlyProGlyValG	137	
692	QY	AAGGAGGGATGAC	CCAGTACACCTC---CGATAGTGACAGCGATGGCTGACCTTGA	748	
137	Db	naArgAlaGlyGlnPro	LeuGlyIleHisArgLeuValGluValAlaGlyAspGlnG	157	
749	QY	AAGAGCTTCAGAATCGCTT	TCGACGAGCGGGAAACAGGACCCACCTGAGAG-----	800	
157	Db	nglnAlaValGlnAlaArg	GlnAlaGlyAlaAspThrAspAlaHisGlnHisGlnAlaVa	177	
801	QY	-----	CCCCCTGAAGGGATCCAGAGTCGCTCGGGAAGACGCGCGG	844	
177	Db	lAlaAlaGlyGlyAlaPro	AlaArgGlnGlnHisArgGlnArgArgAlaGluAl	197	
845	QY	AGAGGGTCCCGCGAGCTGT	GGCTCCGAGGCCAGTGACACTGTGGAGGGCGCTCTCTGC	904	
197	Db	aGlyGlyGlnArgGln-----	-----ArgGlyG	206	
905	QY	CCAGTAAAGCGGAGCC	CAGAACCAT-----CAGGGGGTTGTCTCCAGGCTGGGA	955	
206	Db	yGlnArgAlaGlyIle-Glu	GlnAspAspLeuHisAspGlyAlaAlaGlyAlaGlyG	226	
956	QY	AAGATGACAGAGAGTAGT	TGAGGGAAGCGGCTCAGGACATCAAGATCAGGAGC	1015	
226	Db	lyGluThrAspAspValGly	ArgAlaGlnArgIleAlaArgAspArgLeuGlnAspArgP	246	
1016	QY	CTGAGGACTTGGGCGCG	CCCAAGCCTGAAATGAGGTTCAGACCCCAACGCCCTGTATT	1075	

Db	246	roGlyHisAlaGlnArgGlyAlaAspGlnAlaGlyGluHisProTrpGlnAlaTyra	266
Qy	1076	GCATTTGCCGCCAGCCTCACAAACAGGTTTATGTTGCTGACCGCTGTGAGAAAT	1135
Db	266	laLeu-----AspAspA	270
Qy	1136	GGTTTCATGGCATTTGTGGGCATTTCTGAGCTCGAGGAGGCTTTTGGAAAGGAATG	1195
Db	270	laLeuGlyGlyGlyValLeuAlaLeuAlaGlnAlaGlyGluHisLeuCyArgAlaG	290
Qy	1196	GGGAAGACTATATCTGCCCAAATGCAACCATTTCTGCAAGTGCAGGATGAGACTCATTCAG	1255
Db	290	lnArgGlnAlaAlaGlyProGlnGlnArgArgGlnArgGlnSerAspProArgGluA	310
Qy	1256	AAACGGCGATGATCAGCAGAGCTAAATGCGAGACTGGAGATGCTGATGCGCACCGATTGTA	1315
Db	310	spGlnAlaGlyGlnArgArg---ProTrpProThrGlnArgGlnGlyAlaGlnAlaG	329
Qy	1316	CAAGTATAGGAACAATAAGACAGAAG-----	1341
Db	329	lnGlyValGluThrValHisArgArgGluArgGlyArgLeuThrArgThrMetIysA	349
Qy	1341	-----	1341
Db	349	laGlyProProSerSerIysAlaThrIleProAlaAlaIleSerAlaGlyAlaAlaSerS	369
Qy	1342	-----TCTAGCCAAAGACCAAGGATA-----AAGGTA	1369
Db	369	erArgProArgMetSerAlaProSerSerArgProAlaAlaIleSerAlaGluLysGlyA	389
Qy	1370	GAATTGAGAAAGCTGCAAAATCCAAGTGGCAAG-----AAGAAAC	1408
Db	389	rgileGlnArgTrpSerGlyProAlaSerSerArgAlaArgCysGlyAlaThrArgProT	409
Qy	1409	TCAGATCTCCAGCCT-----GTGATAGAGCGCTGTGTGCTCA-----	1449
Db	409	hrIysAlaIleGlyProAlaArgAlaValAlaAlaProAlaSerSerIleThrAlaArg	429
Qy	1450	-----AAATGTATTGGCCCCGGGTGCTCTCAGCTGGCGCAGCCCG	1489
Db	429	lnAlaSerSerGlnValArgArgArgSerThrProSerAlaArgProIleSerTrpProA	449
Qy	1490	ACTCGGTACTCAGTAATGACTGTGTATCTCAACACGCGCAGCGACCAATGAAGTTTC	1549
Db	449	rgAlaArgAlaLeuSerLeuAlaAlaArgAlaLysAlaAlaAlaSerProAlaProMetA	469
Qy	1550	TAAGCTCAGGT-----AAAGACAGAGCAAGCCATAAGAAAAGCAAGATGAAGC	1603
Db	469	laArgProGlyHisProSerArgAlaLysProLeuProGluSerGluProAlaAlaGlnT	489
Qy	1604	CAGAGAAGCCAGCTTCGGAATGCCGTGCT-----CAGGCAGGTATTAAATCT	1654
Db	489	rpArgIysArgSerLysThrSerGlyGlyAlaPheSerThrMetIleProValSerGluA	509
Qy	1655	CTTCTGTGCACAGACAGCAGCTCCAGAAAAAAGACACACA-----GTGAGGAAGG	1708
Db	509	laSerIleThrAspLysProAlaProAlaSerThrArgArgThrGlyLeuValProValS	529
Qy	1709	CAGTGTGTGCTCCTCGCGGAGTGAACACTCGGGAGAGGAAGCAGCTTGTGAGACAGCA	1768
Db	529	erArgAlaTrpProArgArgAsnThrAlaMetProAlaArgProAlaProArgIysAlaG	549
Qy	1769	CGCCGTG-----TGGCGAGCGATCACAAATTACATGTGAGTAAAGCCAGAAAAGACTGCTG	1825
Db	549	lnProArgTyTrpProMetLeuAlaArgSerAsnSerAlaThrProThrIleThrAlaS	569
Qy	1826	CTCCCTCGCCGTCTACTGTTTATAAATGTATGTATCATCTAGGGGTTGGCCCTCTCTGGACC	1885
Db	569	erAlaAlaProAlaLeuIle-----p	576
Qy	1886	CTCCCGTCTCTTCTGATAGCCATCCCTGGCGCTCTCAGGACTGGGAGTTCAGCTT	1945
Db	576	roSerArg-----ProGlySerAlaSerGlyLeuArgValSerAla-----	589

QY 1946 TGTGTAAAGCTGATCACAGACACCGGCTGCACCATCACCGGGAAGCAGAGCCCATGTCCA 2005
Db 590 Cysile-SerAlaProAlaThrProSerAlaAlaProAlaSerSerAlaThrArgValar 609
QY 2006 G 2006
Db 609 g 609

RESULT 3
US-08-801-308-1
Sequence 1, Application US/08801308
Patent No. 6368790
GENERAL INFORMATION:
APPLICANT: Scott, Robert E.
TITLE OF INVENTION: CDNA ENCODING P2P PROTEINS AND USE OF
TITLE OF INVENTION: P2P CDNA-DERIVED ANTIBODIES AND ANTISENSE REAGENTS IN
TITLE OF INVENTION: DETERMINING THE PROLIFERATIVE POTENTIAL OF NORMAL,
TITLE OF INVENTION: ABNORMAL AND CANCER CELLS IN ANIMALS AND HUMANS
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Weiser & Associates, P.C.
STREET: 230 S. Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/801,308
FILING DATE: 18-FEB-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCES/DOCKET NUMBER: 372,6435P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8393
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1404 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-801-308-1

Alignment Scores:
Pred. No.: 1,53e-05 Length: 1404
Score: 169.00 Matches: 140
Percent Similarity: 32.68% Conservative: 77
Best Local Similarity: 21.08% Mismatches: 238
Query Match: 3.60% Indels: 210
DB: 4 Gaps: 29

US-09-787-016A-1 (1-2610) x US-08-801-308-1 (1-1404)
QY 262 GGTATGAGCAGCAAGCGCAGCCGATGAGGAGCCTAAGCCATCAACCCACC 321
Db 829 GlyProGlnGlnLysValAspGlyAspArgGlnLysSerProArg---SerGluProPr 847
QY 322 AGCAAGAGTTTCAGGAACAACATGGGGTTTCGAAAGGACCACTATCCCAAGCGAGGGCC 381
Db 847 OLeuLysLysAlaLysGlnLysAlaThrLysLysLysLysLysLysLysLysLysLys 867
QY 382 GAGGGGAGCGGAGGCTGACCCATGAGCGCCACCCCGCCAGCAGGAGCTGGCCCTG 441
Db 867 rGlnLysAspGlnValThrGlyThrProArgLysAlaHisSerLysLysLysLysLys 887

QY 442 -----TCCCTGGCGCGCAGTGGGAGG----- 462
Db 887 pThrArgGlnSerGlnProArgThrArgArgSerLysArgThrValProLysThrSe 907
QY 463 -----CAGCCCAAGCGCAGTGCAGCGGTGAGCAGTTCCTGACCAAT----- 504
Db 907 rSerGlnLysSerGlnProValArgThrArgArgProArgSerLysLysLysLysLys 927
QY 505 -----CGCGGCGCGCGCGCAGGAGGAGCATGCTGTCTCCCT 542
Db 927 rLeuIleAlaArgGlnLysAsnGlnArgGlnLysArgLysLysSerValAspLysAspPh 947
QY 543 GAGGAGATTCTGGT-----GAGCCCAACCTCTCTGCCCCCGCCACAGAGCC 584
Db 947 eGluSerSerSerMetLysLysLysValGlnGlyThrGluIleValLysProSerPr 967
QY 585 CGACAGCCTCCGAGCGGCGCAGTGCAGCGCTCTGAGCAGCAGAGCGCCCGCCAGTC 644
Db 967 oLysArgLysMetGlnGlyAspValGlnLysLysLysLysLysLysLysLysLysLys 984
QY 645 TGCTTCCACAGCTGTGAAGGAGCAGCAGCAGCTCTCTGAAAGAGTGAAGAGGGGATGA 704
Db 984 sAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1002
QY 705 CCACGATGACACCTCCGATAGTACACCGATGCGCTGACCTTGAAGAGCTTCAGATCG 764
Db 1002 u-----ThrGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1010
QY 765 CTTTCGCGAAGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 824
Db 1010 aGluAsnAlaSerThrThrLysGluProSerGluLys---LeuGluSerThrSerSerLy 1029
QY 825 CCTG-----CGAAGAGCGCGCGGAGGGGCTCC 854
Db 1029 sLysLysGlnLysValLysGlyLysAlaLysArgLysValAlaGlySerGluLys 1049
QY 855 CGCGAGACTCTG----- 867
Db 1049 rSerSerThrLeuValAspTyrThrSerThrSerThrGlySerProValArgLy 1069
QY 868 -GGCTCCGAGCGCAGTGCAGTGCAGGCGGCTCTGCGCCAGTGAAGAGGAGCGCGAGAA 926
Db 1069 sSerGluGlnLysThrAspThrLysArgThrValLysLysLysLysLysLysLysLys 1089
QY 927 CGAT-----CAGGGGGTGTGTCCAGCGC 950
Db 1089 rAspAsnThrAlaProAlaGluAspValLysLysLysLysLysLysLysLysLysLys 1109
QY 951 TGGGAAAGATCAGCAGAGAGTAAAGTTGGAGGAAAGCGCGCTCAGGACATCAAGATGA 1010
Db 1109 pAspLysAspPheGluSerGluGluLysLysLysLysLysLysLysLysLysLysLys 1124
QY 1011 GGAGCCT---GGAGACTTGGCGCCAGCG-----AAGCC 1040
Db 1124 rGlnProIleGlnSerValGlnLysProSerSerLysLysLysLysLysLysLysLysLys 1144
QY 1041 T-----GAATGAGGGTATGACCCCGCCAGCGCGCTGTATTCAT 1079
Db 1144 cSerAlaThrAlaLysTyrThrGluLysGluSerGluGlnProGluLysLysLysLys 1164
QY 1080 TTGCGCGCAGCTCAGCAACAACAGTTTATGCTGTGTCACCGCTGTGAGATGTT 1139
Db 1164 uProLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1177
QY 1140 TCATGGCGATTGTGGCGCATTTCTAGGCTCGAGGAGGCTTTTGAAGAG-----AA 1193
Db 1177 uArgSerSerLysGlySerAlaSerSerGluLysGlyArgAlaLysAspArgGluHis 1197
QY 1194 TGGGGAAGACTATATCTCCCAAC-----TG 1220
Db 1197 rGlySerGluLysAspAsnProAspLysArgLysSerGlyAlaGlnProAspLysGluSe 1217
QY 1221 CACCATTCTGCAGTGCAGGATGAGACTCATTCAGAAACGCGCAGATCAGCAGGAAGTAA 1280

```

Db      1217 rThValAspArgLeuSerGluGlnGlyHisPhelyThrLeuSerGlnSerSerlysl 1237
QY      1281 ATGCAGACCTGGAGATGCTGATGGCCAGCGATTGTACAGTATAGGAACAATAGAGCAGAA 1340
Db      1237 uThrArg-----Th 1240
QY      1341 GTCTAGCGAAGACCAGGATTAAGGTTAGATTGAGAAAGCTGCAATCCAGTGGCAA 1400
Db      1240 rSerGluLysHisGluSerValArgGlySerSerAsnLysAspPheThrProGlyArgAs 1260
QY      1401 GAAGAAACTCAAGATCTTCCAGCCTGTGATAGAGCGCTGGTGCCTCAAAATGATTGG 1460
Db      1260 pLysLysValAsp----- 1264
QY      1461 CCCCAGGTGCTGTACGTGGCGCAGCCGACTCGGTGTACTGCAAGTAATGACTGTATCT 1520
Db      1265 -----TyrAspSerArgAspTyr-SerSe 1272
QY      1521 CAACACCGCGCAGCAGCAATGAAGTTTCTAAGCTCAGGTAAAGACAGACAGCAAGCC 1580
Db      1272 rSerLysA-gAAspGluArgGlyGluLeuAlaAArgArgLysAspSerProProArgGl 1292
QY      1581 TAAAGAA-----AAGATGAAGATGAAGCCAGAGAACCCAGTCTTCCGAA 1625
Db      1292 yLysGluSerLeuSerGlyGlnLysSerLysLysLeuAArgGluAArg---AspLeuProLy 1311
QY      1626 ATCCGCTGCTCAGCAGGTATTAAATCTCTTCTGTG-----CACAGAGACCACTCC 1679
Db      1311 sLysGlyAlaGluSerLysLysSerAsnSerSerProA-gAspLysLysProHisAs 1331
QY      1680 AGAAAAAAGAACACACAGTGAAGAGGAGGAGTGTGTGTCTCTCCGCGGAGTGAAGCACT 1739
Db      1331 pHisLysAlaProTyrGluThrLysArg----- 1340
QY      1740 CGGGAAGGAAGCAGCTTGTGAGAGCAGCAGCGCGTGTGGGAGGAGATCACAATTACAA 1799
Db      1341 -----ProCysGluGluThrLysPro-----ValAspLysAsn---Se 1352
QY      1800 TGCAGTAAAGCCAGAAAGACTGCTGCTCCTCG-----CCGTCACCTGTGTA 1847
Db      1352 rGlyLysGluArgGluLysHisAlaAlaGluAlaAArgAsnGlyLysSerSerGlyAl 1372
QY      1848 TAAATGTATGTATACACTAGGGTTGGCTTCTCGACCCCTCCGTTCTTCTGATAGC 1907
Db      1372 aAsnCysHisValTyrLeuThrArgGlnThrLeu---ProTyrArgArgSerTyrLeuLe 1391
QY      1908 CATCCCTCG 1917
Db      1391 uGlyArgTyr 1394

```

RESULT 4

```

US-09-252-991A-18387
; Sequence 18387, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18387
; LENGTH: 1075
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18387

```

```

Alignment Scores:
Pred. No.: 3,24e-05 Length: 1075
Score: 164.50 Matches: 126
Percent Similarity: 34.53% Conservative: 76
Best Local Similarity: 21.54% Mismatches: 228
Query Match: 3.51% Indels: 155
DB: 4 Gaps: 26

US-09-787-016A-1 (1-2610) x US-09-252-991A-18387 (1-1075)

QY      283 CCGACCAATGAGGAGGACCTTAAGCCATCAACCCACGACCAAGAGATTTCAGGAAACA 342
Db      553 ProGlnHisThrGluAlaProValGluProAlaLysProMetProGluProSerLeuPhe 572
QY      343 TGGGATTTTCGAGGACCACTATCCCAAGCGAGAGGCGCAGGAGGACGCGGAGGCTGAC 402
Db      573 GlnGlyLeuValLysSerLeuValGlyLeuPheAlaGlyLysAspGlnProAlaLys 592
QY      403 CCACCTGGAG-----CCGCCACCCCCACGACGAGCTGGGCTCTCCTCGCGGCGAGT 456
Db      593 ProAlaGluThrSerLysProAlaAlaGluArgGlnThrArgGlnAspGluArgAsn 612
QY      457 GGGAGGAGCCCGACGAGCTCCGAGGCGAGCGTGGAGAGCGCTTCTGAGACGACGAGCGGC 636
Db      646 -----GluArgAlaGluArgProAsnArgGluGluAArgSerGluArgArgGlu 662
QY      637 CCCAGTCTGCTCCACAGCTGTGAAGGAACGACGACGCTCTCTGAAAGAGGTGAAGGA 696
Db      663 GluArgAla-----GluArgProAlaAArgGluGluArg----- 673
QY      697 GGGGATGACCAAGATGACACCTCCGATAGTGCACCGATGGCTCTCCTTGAAGAGCTT 756
Db      673 ----- 673
QY      757 CAGATCCGCTTCGAGGAGCGGGAACAGAGCCCTGAGAGG-----CCC 804
Db      674 GlnProArgGluGlyArgGluGluArgAlaGluArgThrProArgGluGluArgGlnPro 693
QY      805 CTGAAGGGATCCAGAGTCCG-----CTGGGGAAGAACCGCGGAGAGGCTCCCGCGAG 861
Db      694 ArgGluGlyArgGluGlyArgGluGluArgSerGluArgArgArgGluGluArgAlaGlu 713
QY      862 ACTGTGGCTCCGAGGCCAGTGCACACTGTGAGGGGCTCTCCGCCAGTGAAGCAGGAGCC 921
Db      714 ArgProAlaAArgGluGluArgGlnProArgGluGlyArgGluGluArgValGluArgPro 733
QY      922 GAGAAGCATCAGGGGTTGTGTCCAGGCTGGGAAGATGACAGAGAGAGTAAG----- 975
Db      734 AlaArgGluGlu-----ArgGlnProArgGluAspArgGlnAlaArgAspAla 749
QY      976 -----TTGGAGGGAAGGCG-----GCTCAGACATCAAGAT 1008
Db      750 AlaAlaLeuGluAlaGluAlaLeuProAsnAspGluSerLeuGluGlnAspGluGlnAsp 769
QY      1009 GAGGAGCTCGAGACTTGGCGCGACCGGAGCTGATGTGAGGTTTACGACCCCAAGCC 1068
Db      770 AspThrAspGlyGlu-----ArgProArgArgArgSerArgGlyGln----- 783
QY      1069 CTGTATTGCATTTCGCCGAGCTCACAACAACAGGTTTATGATTGTGTGACCGCTGT 1128
Db      784 -----ArgArgArgSerAsnArgArg-----GluArgGln 793
QY      1129 GAAGAATGTTTCATGGCGATGTGTGGCATTTCTGAGGCTTCGAGGAGGCTTTTGGAA 1188
Db      794 ArgGlu---ValSerGlyGluValGluGly---SerGluAlaThr----- 806

```

```

QY 1189 AGGAATGGGAGAGACTATATCTGCCCCAAACTGCACCATTT----- 1227
Db 807 -----AspAlaAlaAlaProLeuAsnThrValAlaAlaAlaAlaAlaGly 823
QY 1228 CTGCAAGTCAGGATGAG-----ACTCATTC 1254
Db 824 IleAlaValAlaSerGluAlaValGluAlaAsnValGluGlnAlaProAlaThrSer 843
QY 1255 GAAACGGCAGATCAGCAGGAAGCTAAATGAGACCTGGAGATGCTGATGGCACCATTGT 1314
Db 844 GluAlaAlaSerGluThrThrAla-----SerAspGluThrAspAla 857
QY 1315 ACAAGTATAGGACATAGACAGAGAGTCTAGCGAGACCAAGGGATTAAGGGTAGA--- 1371
Db 858 SerThrSerGluAlaValGluThrGlnAspAlaAspSerGluAlaAsnThrGlyGluThr 877
QY 1372 -----ATTGAGAAGCTGCAAAATCCAAAGTGGCGCAAGAGAAA 1407
Db 878 AlaAspIleGluAlaProValThrValSerValValArgAspGluAlaAspGlnSerThr 897
QY 1408 CTCAGATCTTCAGCTGTAGAGCGCTGTGCTGCTCAAAATGTATT----- 1458
Db 898 LeuLeuValAlaGlnAlaThrGluGluAlaProPheAlaSerGluSerValGluSerArg 917
QY 1459 -----GGCCCC 1464
Db 918 GluAspAlaGluSerAlaValGlnProAlaThrGluAlaAlaGluGluValAlaAlaPro 937
QY 1465 GGGTGTGTGTCAGTGGCGAGCCGACTGGTGTACTGCGAGTAATGACTGTATCTCCANA 1524
Db 938 ValProValGluValAlaAlaProSerGluProAlaAlaThrGluGluProThrProAla 957
QY 1525 CAGCCGCGCA-----GGCAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAG 1572
Db 958 IleAlaAlaValProAlaAlaThrGlyArgAlaLeuAsnAspProArgGluLysArg 977
QY 1573 CCAAAGCCTAAAGAAAGATGAAGTAAAGCCAGAGAGCCCGCTTCGGAAATCGGT 1632
Db 978 ArgLeuGlnArgGluAlaGluArgLeuAlaAlaArgGluAlaAlaAlaGluAlaAla 997
QY 1633 GCTCAGCGAGTATTAAATCTCTCTGTGCAAGAGACAGCTCCAGAAAAAAGAG 1692
Db 998 AlaGlnAlaAla-----ProAlaValGluGluIleProAlaValAlaSerGluGlu 1014
QY 1693 ACCACAGTGAAGAGCAGTGGTGGTCCCTGCGCGAGTGAACACTGGGAAAGGAAGA 1752
Db 1015 AlaSerAlaGlnGluGlu-----ProAlaAlaProGlnAla-----GluGlu 1028
QY 1753 GCTGTGAGACGACCGCGCTGTGGCGAGCGGATCACATACATGAGTAAAGCCA 1812
Db 1029 IleThrGlnAlaAspValProSerGlnAla---AspGluAlaGlnGluAlaValGlnAla 1047
QY 1813 GAAAAGACTCTGCT 1827
Db 1048 GluProGluAlaSer 1052

```

RESULT 5

```

US-09-252-991A-27005
; Sequence 27005, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Merck J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: 1999-02-18
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```

; SEQ ID NO 27005
; LENGTH: 1427
; TYPE: FRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27005

```

```

Alignment Scores:
Pred. No.: 3,7e-05 Length: 1427
Score: 164.50 Matches: 111
Percent Similarity: 36.18% Conservative: 50
Best local Similarity: 24.94% Mismatches: 128
Query Match: 3.51% Indels: 156
DB: 4 Gaps: 25

```

US-09-787-016A-1 (1-2610) x US-09-252-991A-27005 (1-1427)

```

QY 273 CAAGAGCGACCCGACCAATGAGGAGGACCTTAAGCCATCAACCCACCCAGCAAGAGTT 332
Db 770 ArgArgGlnProGluPro---GlyAlaValProGlyArgAlaAlaHisArg----- 785
QY 333 CAGGAAACATGGGGTTTTCGAGGACCACTATCCCAAGCGAGAGGGCGGAGGGAGCG 392
Db 786 -----GlyAspProTyrLeuArgProArgHisArgArgGlyArg 798
QY 393 GGAGCTGACCCACTGGAGCGCGCCACCCACAGCAGCAGCTGGGCTGTCTCCGCGCG 452
Db 799 AlaArgGlyHisAlaGlyAlaAla---ProHisGlnProAlaGly-----AlaAla 814
QY 453 CAGTGGGAGGAGCGACCAAGCGCACTGAGCGCGTGGAGCAGTTCCTGACANTGCGCGCG 512
Db 815 LeuLeuGlyAlaAlaGluAspHis---SerArgTyrArgAlaAlaAspHisProAlaGly 833
QY 513 CCG-----CGCAGGAGGAGCGCTGCC----- 533
Db 834 AlaAspArgLeuSerLeuArgArgArgArgHisAlaHisAlaArgThrAlaLeuHis 853
QY 534 -----TGCTCCCT-----GGAGGATTC 551
Db 854 ProAlaAlaValArgArgLeuProAlaLeuAlaArgGluGlnArgThrLeuGlyGly--- 872
QY 552 TGGTGAAGCCCACTCTCCCGCCACAGCGCAGCAGCAGCTCCGAGGGCAGCGTGA 611
Db 873 -----AlaGluLeuValProHisHisArgArgGlyArgHisArgProThrLeu 890
QY 612 AAGCGCTTCTGAGACCAAGCGCGCCCGCTCTCCACAGCTGTGAAGAAAGCA--- 669
Db 891 LysArg-----ArgArgPro-SerArgSerValSerGlnArgArgAl 905
QY 670 -----CCAGCTCTCTTGAAGAGTGAAGAGGGGATGACCCAGATGACACCTC 719
Db 905 aLeuAlaThrProArgThrGluGlnArgGlnHisGlyGlyArgAspAspGluHisArg 925
QY 720 CGATAGTCAGCGATGCTGCTGACCTTGAAGAGCTTCAGATCGCTTCGAGGAAGCG 779
Db 925 gGluArgProIleGluGly-----GlnArgLeuProAlaValVa 938
QY 780 GGAACAGAGCGCCACTGAGAGGCGCTTGAAGGGATCCAGAGTCGCTCGGAAAGAGCG 839
Db 938 lGluGlnGluProGlyHis-----GlnAr 946
QY 840 CCGGAGAGAGGGTCCCGCCGAGACTGTGGCTCCGAGGCCAGTGAC-----ACTGT 890
Db 946 gArgGluAspArgProAlaAlaThrAspAlaAspProProAlaAspProGlyGlyThrHi 966
QY 891 GGAGGCGCTCTCCCGCAAGTAAGCAGGAGCCCGAGAACGATCAGGGGGTTGTGTCCAGCG 950
Db 966 sArgSerValValGluAlaArgArg-----GlnGlyIle-----GlnAl 979
QY 951 TGGG-----AAGATGACAGAGAGTAAAGTTCGAGGAGAAAGCG--- 990
Db 979 aGlyHisAlaGlyValGlyAlaLysProAspGlnSerGlnArgGlnGlyArgAlaGl 999
QY 991 -----GCTCAGGACATCAAAGATGAGGAGCTGGAGAC----- 1023

```

```

Db      999 yclnPrGlyAlaGlnArgGluGlnSerGlyGlyLysGlnHisGluGlnAr 1019
QY      1024 -----TTGGCCCGACCAAGCCTGAATGTGAGGTTTACGA 1058
Db      1019 gclnHisProPheGlnAlaLysAlaArgGlyLysProAlaGluGlnGlyThrGluAs 1039
QY      1059 CCCCAGGCCCTGATTGCATTTGCCCGCAGCCTCACAAACAGGTTTATGATTTGCTG 1118
Db      1039 pPro-----1040
QY      1119 TGACCGCTGTGAAGAATGGTTTCATGCGCATTTGTGGGCATTTCTGAGCTCGAGG-- 1176
Db      1041 -AlaAspAlaGlnGlnGlyAlaLysAlaArgGlyLysProAlaGlnGlyGlyAs 1060
QY      1177 -AGCTTTTGAAGAAGAAAGAGAGACTATATCTGCCCAACTGCACCATTTCTGCAAGT 1235
Db      1060 pGlnGlnLeuArgProAlaHisAsp-----GluVa 1071
QY      1236 GCAGGATGAGACTCATTTCAAGAAACGCCAGATCAGCAGGAAGCTAAATGAGACCTGGAGA 1295
Db      1071 lGluAlaGluHisGlnGlu-----LysGlnArgProGluGln 1084
QY      1296 TGCTGATGCCACCGATTGTACAGTATAGTAAAGGATTAAGTGAAGAGCTGCAATCC 1391
Db      1084 nGlnArgGly-----ThrSerLeuGlyAsnAlaLysGlnProGlyGluAlaAlaTh 1101
QY      1339 -----AAGTCTAGCAGAGACCAAGGATTAAGGATTAAGTGAAGAGCTGCAATCC 1391
Db      1101 rValLeuArgSerAlaThrGlyArgAlaArgGlnPheArgGluArgAlaAlaGlnAl 1121
QY      1392 AAGTGGCAAGAG 1404
Db      1121 aProGlyGlnArg 1125

RESULT 6
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauk, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta B.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-769-309A-5

```

Alignment Scores:

```

Pred. No.: 4.97e-05 Length: 1780
Score: 163.50 Matches: 136
Percent Similarity: 30.71% Conservative: 79
Best Local Similarity: 19.43% Mismatches: 224
Query Match: 3.49% Indels: 261
DB: 1 Gaps: 29

```

US-09-787-016A-1 (1-2610) x US-08-769-309A-5 (1-1780)

```

QY      226 TCAGGAAAGTGTCCAAAGCTTTCAGTCTTCGAGCAGGTATGAGCAGCAAGCGCACCG 285
Db      257 SerGlyGlnAlaValGluGlu-----CysLysGluGluGlyGluGlu 270
QY      286 AGCATATGAGGAGCACCT-----AAGCCATCAAAACCCACAGC----- 324
Db      271 LysGlnGluLysGluProSerLysSerAlaGluSerProThrSerProValThrSerGlu 290
QY      325 -----AAAGATTTCAGAAACATCG-----GGTTTTCGAGGACCAT 363
Db      291 ThrGlySerThrPheLysLysPhePheThrGlnGlyTrpAlaGlyTrpArgLysLysThr 310
QY      363 ----- 363
Db      311 SerPheArgLysProLysGluAspGluValGluAlaSerGluLysLysGluGlnGlu 330
QY      364 -----ATCGCCAAAGGAGGAGCGCGCA 384
Db      331 ProGluLysValAspThrGluGluAspGlyLysAlaGluValAlaSerGluLysLeuThr 350
QY      385 GGGGACGCGGAGGCTGACCCACTGAGCGCGCA----- 417
Db      351 AlaSerGluGlnAlaHisProGlnGluProAlaGluSerAlaHisGluProArgLeuSer 370
QY      418 -----CCCCCAGCAGCAGCTGGGCGCTGTCCCTGGCGGC 453
Db      371 AlaGluTyrgLysValGluLeuProSerGluGluGlnVal----- 384
QY      454 AGTGGAGGCGAG-----CCCAAGCGCAGCTGAGCGGTGGAG 489
Db      385 SerGlySerGlnGlyProSerGluGluLysProAlaProLeuAlaThrGluValPheAsp 404
QY      490 CAGTTCTCTGACCATTCGCGCGCGCGCGCAGGAGCATCGCTCTCTCCCTGGAGGAT 549
Db      405 GluLysIleGluValHisGlnGluValValAlaGluValHisValSerThrValGlu 424
QY      550 TCTGTGAGCCCGACGTCTCTGCCCCCGCCACAGACGCC-----GAGACAGCCTCCGAGGCGAG 606
Db      425 -----GluArgThrGluGluGlnLysThrGluValGluGluThrAlaGlySerValPro 442
QY      607 GTGGAAGCGCTTCGAGACCAAGCGCGCCCTGCTCTCCAGCTGTGAAGAA 666
Db      443 AlaGluGluLeuValGlyMetAspAlaGluProGlnGluAlaGluProAla-----LysGlu 461
QY      667 CGACAGCGCTCTCTGAAAGGTGAAGAGGAGATGACACGATGACACCTCCGATAGT 726
Db      462 LeuValLysLeuLysGluThrCysValSerGlyGluAspProThrGlnGlyAlaAspLeu 481
QY      727 GACAGCGATGCGCTGACCTTGAAGAG----- 753
Db      482 SerProAspGluLysValLeuSerLysProProGluGlyValValSerGluValGluMet 501
QY      754 CTTTCAAGATCGCTTTCGAGGAAGCGGGAACAGAGAGCCC-----ACTGAG 798
Db      502 LeuSerSerGlnGluArgMetLysValGlnGlySerProLeuLysLysLeuPheThrSer 521
QY      799 AGGCCCTTGAAGAGGATCCAGATGCTCCCTCGGAAGAGAGCGCGGAGGAGCGGT----- 852
Db      522 ThrGlyLeuLysLysLeuSerGlyLysGlnLysGlyLysArgGlyGlyAspGlu 541

```

853 QY -----CCGCGCAGAGACTGTGGGCTCCGAGCGCCAGTGCACACT 889
542 Db GUserGlyGluHisThrGlnValProAlaAspSerProAspSerGlnGluGlnLys 561
889 QY GTGGAGGGGGTCTGCGCCAGTAAGCAGGAGCCGAG----- 924
562 Db GlyGluSerSerAlaSerSerProGluGluProGluGluLeuThrCysLeuGluLysGly 581
925 QY -----AAGCATCAGGGGGTGTGCTCCAGGCTGGGAAGATGACAGAGAGT 972
582 Db LeuAlaGluValGlnGlnAspGlyGluAlaGluGlyAlaThrSerAspGlyGluLys 601
973 QY AAGTTGGAGGA----- 984
602 Db LysArgGluGlyValThrProTrpAlaSerPheLysLysMetValThrProLysLysArg 621
984 QY ----- 984
622 Db ValArgArgProSerGluSerAspLysGluAspGluLeuAspLysValLysSerAlaThr 641
985 QY -----AAGCGGCTCAGGACATCAAGATGAGGAGCTCGAGACTTGGC 1029
642 Db LeuSerSerThrGluSerThrAlaSerGluMetGlnGluMetLysGlySerValGlu 661
1030 QY CGACCAAGCCTGAATGTGAGGTTACGACCCCAAGCCCTGTATTGTCATTTGCCGCCAG 1089
662 Db GluProLysProGlu-----Glu 667
1090 QY CTTCAACACACAGGTTTATGATTGCTGTGACCGCTGTGAAGATGTTTCATGCGAT 1149
668 Db ProLysArgLys-----ValAspThrSerValSerTrpGluAlaLeuIle 682
1150 QY TGTGTGGCATTTCTCAGGCTCGAGGAGCTTTGGAAAGAAATGGGAAGCATATATC 1209
683 Db CysValGlySerSerLysLysArgAlaArg-----ArgArgSerSerSerAsp 698
1210 QY TGCCCAAACTGCACCATTTGCAAGTGCAGATGAGACTCATTCAGAAACCGCGATGATC 1269
699 Db -----GluGluGlyGlyProLysAlaMetGlyGlyAspHis 710
1270 QY CAGAAAGCTAAATGAGACCTGGAGATGCTGAT---GGCACCATTGTCAAGTATAGA 1326
711 Db GlnLysAlaAspGluAlaGlyLysAspLysGluThrGlyThrAspGlyLysLeuAlaGly 730
1327 QY AACTAGACAGAGTCTACGAGAACCAAGGGATAGAGTGAATGAGAAAGCTGCA 1386
731 Db Ser-----GlnGluHisAspProGlyGlnGly---SerSerSerProGluGlnAlaGly 747
1387 QY ATCCAAAGTGGCAAGAAACTCAAGATCTCCAGCCTGTGATAGGCGCCTGTGCC 1446
748 Db SerProThr-----GluGlyGluGlyVal 755
1447 QY TCAAAATGATATGCCCCGGGTGTGTACGTGGCGCCGACTCGGTGTACTCAGT 1506
756 Db Ser----- 756
1507 QY ATGACTGTATCTCAACACGCGCGCGCAACATGAGTTTCTAAGCTCAGTAAAGAA 1566
757 Db -----ThrTrpGluSerPheLysArgLeuValThr----- 766
1567 QY CAGAAGCAAAAGCTAAAGAAAGATGAAGATGAAGCAGAGAGCCAGCTCTTCGAAA 1626
767 Db -----ProArgLysLysSerLysSerLysLeuGluGluLysSerGluAspSer 782
1627 QY TGGGTGCTCAGGAGGTATTAATCTCTTGTGTGCACAAAGAGACAGCTCCAGAAAA 1686
783 Db Ile-----AlaGlySerGlyValGluHisSerThrProAspThrGluProGlyLysGluGlu 801
1687 QY AAGAGACCACTGAGAGAGGCGTGTGTCCTCGCGGAGTGAAGCACTCGGGAAG 1746
802 Db SerTrpValSerIleLysPheIleProGlyArgArgLysLysArgProAspGlyLys 821

1747 QY -----CAAGCAGCTTGTGAGACGACGCGCTGTGGCGGCGGATCAAT----- 1794
822 Db GlnGluGlnAlaProValGluAspAlaGlyProThrGlyAlaAsnGluAspSerAsp 841
1795 QY -----TACATCAGCTAAAGCCAGCAAAAGACTGTGCTGCT 1827
842 Db ValProAlaValValProLeuSerGluTyraAspAlaValGluArgGluLysMetGluAla 861
RESULT 7
US-08-994-570-5
; Sequence 5, Application US/08994570
; Patent No. 6090929
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauk, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/994,570
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: NO. 6090929and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-994-570-5
Alignment Scores:
Pred. No.: 4,978-05 Length: 1780
Score: 163.50 Matches: 136
Percent Similarity: 30.71% Conservative: 79
Best Local Similarity: 19.43% Mismatches: 224
Query Match: 3.49% Indels: 261
DB: 3 Gaps: 29
US-09-787-016A-1 (1-2610) x US-08-994-570-5 (1-1780)
QY 226 TCAGGAAAAGTGTCAAGCTTTCAGTGTGAGCAGGATATGACGACAAAGGCGACCG 285
Db 257 SerGlyGlnAlaValGluGlu-----CysLysGlnGluGlyGluGlu 270
QY 286 AGCAATGAGGAGGACCT-----AAGGCCATCAACCCACCAGC----- 324
Db 271 LysGlnGluLysGluProSerLysSerAlaGluSerProThrSerProValThrSerGlu 290
QY 325 -----AAAGAGTTCAGAGAAAATGG-----GGTTTCGAGGACCACT 363
Db 291 ThrGlySerThrPheLysLysPheThrGlnGlyTrpAlaGlyTrpArgLysLysThr 310
QY 363 ----- 363

```

Db      311 SerPheArgLysProLysGluAspGluValGluAlaSerGluLysLysLysGluGlnGlu 330
QY      364 -----ATGCCAAGCCAGAGGGCGCA 384
Db      331 ProGluLysValAspThrGluGluAspGlyLysAlaGluValAlaSerGluLysLeuThr 350
QY      385 GGGGACGGAGGCTGACCCACTGGAGCCGCCA-----417
Db      351 AlaSerGluGlnAlaHisProGlnGluProAlaGluSerAlaHisGluProArgLeuSer 370
QY      418 -----CCCCACACAGCAGCTGGGCCCTGTCCTCGCGCGC 453
Db      371 AlaGluThrGluLysValGluLeuProSerGluGluGlnVal-----384
QY      454 AGTGGAGGCGAG-----CCCAAGCGCACTGAGCGCGTGGAG 489
Db      385 SerGlySerGlnGlyProSerGluGluLysProAlaProLeuAlaThrGluValPheAsp 404
QY      490 CAGTTCCTGACCATTCGGCGCGCGCGGAGGAGGAGGATGCTGTCCTCCCTGGAGGAT 549
Db      405 GluLysIleGluValHisGlnGluGluValAlaGluValHisValSerThrValGlu 424
QY      550 TCTGGTGGCCACGCTCTGCCCGCGCCACAGAGCC---GAGACAGCTCCCGAGGCGAGC 606
Db      425 -----GluArgThrGluGluGlnLysThrGluValGluGlnThrAlaGlySerValPro 442
QY      607 GTGGAAGCGCTTCTGAGACCAAGCGCGCCAGTCTGCTCCACAGCTGTGAAGAA 666
Db      443 AlaGluGluLeuValGlyMetAlaGluProGlnGluAlaGluProAla---LysGlu 461
QY      667 CGACCAAGCTCTTCTGAAAGGTGAAGAGGAGGATGACCAACGATCACCTCCGATAGT 726
Db      462 LeuValLysLeuLysGluThrCysValSerGlyGluAspProThrGlnGlyAlaAspLeu 481
QY      727 GACAGCGATGGCTGACCTTGAAGAG-----753
Db      482 SerProAspGluLysValLeuSerLysProGluGlyValValSerGluValGluMet 501
QY      754 CTTGAGATCGCTTCGACGAAGCGGACAGAGCGCC-----ACTGAG 798
Db      502 LeuSerSerGlnGluArgMetLysValGlnGlySerProLeuLysLysLeuPheThrSer 521
QY      799 AGGCCCTGAAAGGATCCAGATCGCTCGGGAAGAGCGCGGAGGAGGT-----852
Db      522 ThrGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyAspGlu 541
QY      853 -----CCGCGAGACTCTGGCTCCGAGGCGGAGTGCACACT 888
Db      542 GluSerGlyGluHisThrGlnValProAlaAspSerProAspSerGlnGluGluLys 561
QY      889 GTGGAGGCGCTCTGCCCACTAAGCAGAGCGCGAG-----924
Db      562 GlyGluSerSerAlaSerProGluGluProGluGluIleThrCysLeuGluLysGly 581
QY      925 -----AACGATCAGGGGTGTGTCACGAGCTGGGAAGATGACAGAGAGT 972
Db      582 LeuAlaGluValGlnGlnAspGlyGluAlaGluGluGlyAlaThrSerAspGlyGluLys 601
QY      973 AAGTTGGAGGCA-----984
Db      602 LysArgGluGlyValThrProThrAlaSerPheLysLysMetValThrProLysLysArg 621
QY      984 -----984
Db      622 ValArgProSerGluSerAspLysGluAspGluLeuAspLysValLysSerAlaThr 641
QY      985 -----AAGGCGCTCAGACATCAAGATCAGGAGCTGAGACTGGGC 1029
Db      642 LeuSerSerThrGluSerThrAlaSerGluMetGlnGluMetLysGlySerValGlu 661
QY      1030 CGACCGAGGCTGAATGTAGGGTTACAGCCCAACCGCCCTGTATTGATTTCCGCCAG 1089

```

```

Db      662 GluProLysProGlu-----Glu 667
QY      1090 CUTCACAACAACAGGTTTATGATTGCTGTGACCCCTGTGAAGATGTTTATGCGCAT 1149
Db      668 ProLysAlaGlyS-----ValAspThrSerValSerTrpGluAlaLeuIle 682
QY      1150 TGTGTGGGCTNTTCTGAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAGACTATATC 1209
Db      693 CysValGlySerSerLysLysArgAlaArg-----ArgArgSerSerSerAsp 698
QY      1210 TGCCCAAACTGCACCATTTCTCAGTCGAGGATGAGACTCATTTCAAGAACCGCAGATCAG 1269
Db      699 -----GluGluGlyGlyProLysAlaMetGlyLysAspHis 710
QY      1270 CAGGAAGCTTAATGAGACCTGGAGATGCTGAT---GGCAGGATGTACAGATAGGA 1326
Db      711 GlnLysAlaAspGluAlaGlyLysAspLysGluThrGlyThrAspGlyLysLeuAlaGly 730
QY      1327 ACAATAGAGCAGAACTCTAGCGAGACCAAGGATAAAGGGTAGAATTCAGAAAGCTGCA 1386
Db      731 Ser-----GlnGluHisAspProGlyGlnGly---SerSerSerProGluGlnAlaGly 747
QY      1387 AATCAAGTGGCAGAGAACTCAAGATCTTCCAGCTGTGATAGAGCGCTGTGCTGCC 1446
Db      748 SerProThr-----GluGlyGluGlyVal 755
QY      1447 TCAAAATGATTGGCCCGGGTGTCTCACGTGGCGCAGCCGACTCGGTGTACTGCAGT 1506
Db      756 Ser-----756
QY      1507 AATGACTGTATCTCTCAACACCGCGAGCGACAATGAAGTTTCTTAAGCTCAGGTAAAGAA 1566
Db      757 -----ThrTrpGluSerPheLysArgLeuValThr 766
QY      1567 CAGAACCCAAAGCCTTAAGAAAGATGAAGTGAAGCAGAGAGCCCGAGTCTCCGAAA 1626
Db      767 -----ProArgLysSerLysSerLysLeuGluGluLysSerGluAspSer 782
QY      1627 TCGCGTGTCTCAGCGAGTATTAAATCTCTCTGTGCACAGAGACCACTCCAGAAAAA 1686
Db      783 Ile---AlaGlySerGlyValGluHisSerThrProAspThrGluProGlyLysGluGlu 801
QY      1687 AAGAGACCACTGAGAGAGGAGCAGTGGTGGTCCCTGCGCGAGTGAAGCACTCGGGAAG 1746
Db      802 SerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgProAspGlyLys 821
QY      1747 -----GAAGCAGCTGTGAGAGCAGCAGCGCTGTGGCGAGCATCAAT-----1794
Db      822 GlnGluGlnAlaProValGluAspAlaGlyProThrGlyAlaAsnGluAspSerAsp 841
QY      1795 -----TCAATGCAAGTAAAGCCAGAAAGACTGTGCT 1827
Db      842 ValProAlaValValProLeuSerGluTrpAspAlaValGluArgGluLysMetGluAla 861

RESULT 8
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; TITLE OF INVENTION: Neuroepithelial Tumors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA: US/07/853,913
 FILING DATE: 19920319
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA: US 07/660,412
 FILING DATE: 22-FEB-1991
 APPLICATION NUMBER: US 07/603,803
 FILING DATE: 25-OCT-1990
 PRIOR APPLICATION DATA: US 07/201,762
 FILING DATE: 02-JUN-1988
 APPLICATION NUMBER: US 07/180,548
 APPLICATION DATA: US 07/180,548
 ATTORNEY/AGENT INFORMATION:
 NAME: Granahan, Patricia
 REGISTRATION NUMBER: 32,227
 REFERENCE/DOCKET NUMBER: MIT-4641AAAA
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-861-6240
 TELEFAX: 617-861-9540
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1805 amino acids
 TYPE: AMINO ACID
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-853-913-2

Alignment Scores:
 Pred. No.: 6,07e-05
 Score: 162.50
 Percent Similarity: 34.81%
 Best Local Similarity: 20.86%
 Query Match: 3.47%
 Indels: 255
 Gaps: 40

US-09-787-016A-1 (1-2610) x US-07-853-913-2 (1-1805)

QY 3 CGCTGGCGCGCCGCTCCGCGGCGTTCGGGAAATGGCTCGAGACCTAGAGCGCT 62
 DB 382 ArgThrProThrLeuAlaSerThrProIleProPheSerGluAlaProCysPro 401
 QY 63 ---CGGAGCTTACTCCACGGGAACAGCCTCTAGATAATCTGAGTTGTTGAAATACGAA 119
 DB 402 AsnAlaGluValArgAlaGluGluValProLeuSerLeuLeuGlnThrGlnAlaProGlu 421
 QY 120 GCCTGTGTAC---TCGTGAACAGTGGCTGACACAGAGTGTCTGTG----- 161
 DB 422 ProLeuTrpLeuAlaThrValProSerSerAlaIleLeuProGluLeuGluGlu 441
 QY 162 -----AGCTTGGCTGTCTGCTG-----AGCTTGGCTGTCTGCTG 179
 DB 442 ProGlyGlyLysGlnGlnGlyHisPheProAspLeuThrSerLeuAlaThrAsnLeu 461
 QY 180 GACCCA-----GAGTTTCGTCTGCCAGGTTTTT 209
 DB 462 AsnProHisHisProThrLeuGluAlaLysAspGlyCysSerGluSerArgValSer 481
 QY 210 GGTGTATTT----- 219
 DB 482 Ser-IlePheGlnGluAspGluGlyGlnIleTrpGluLeuValGluLysGluAlaAsp11 501
 QY 220 -----AGATTTCAGGAAAGTGTCCAGCTTTCAGTGTTCGAGCAGGTATGACGA 272
 DB 501 eGluValLysValGluAsnSerSerAlaGlnLys---ThrGlnGluSerGlyLeuAsp-- 519
 QY 273 CAAGGGGACCGAGCAATGAGGAGGACCTAAGGCCATCAAAACCCACGACGAAAGAGTT 332

DB 520 -----ThrGluGluThrGlnAspSerGlnGlyProLeuGlnLysGluTh 534
 QY 333 CAGGAAACATGGGT-----TTTGAAGGACCACTATCGCAAGCGAGAGCGGC 383
 DB 534 rleuylsalaleuGlyGluGluProLeuMetSerLeuLysileGlnAsnTyTrGluThAl 554
 QY 384 AGGG-----GAGCGGAGGCTGACCCACTGCGAGCGCGCCACC 419
 DB 554 adlyLysGluAsnCysAsnSerSerThrGlyHisLeuGlyThrLeuGluGlyProGI 574
 QY 420 CCCACAGCAGCAGCTGGGCGCTG---TCCTCTGGGCGCAGT----- 456
 DB 574 uLysGluLysGlnIleProLeuLysSerLeuGluGluLysAsnValGluSerGluLysTh 594
 QY 457 -----GGAGGCGAGCCCAAGCGCACTGA 479
 DB 594 rleuGluAsnGlyValProValLeuSerGluLeuLysGluAspThrArgThrGI 614
 QY 480 G-----CGCTGGAGCAGTCTCTGACCAITTCGCGG--- 510
 DB 614 uAspGlnGluLeuMetSerProLysGlyThrLeuLysArgPheSerSerLeuLysGlu 634
 QY 511 -----CGCGCGCGGCGG---AGGAGCATGCTGTCTCCCT 542
 DB 634 uSerGlnGluValValArgProSerLysGluGlyAsnLeuGluSerTrpThrAlaPheLy 654
 QY 543 CGAGGATTCTGTGAGCCCGCTCTGCGCCGACAGCAGCGGAGCAGCTCCGAGG 602
 DB 654 sGluGluSerGlnHisProLeuGlyPheProGlyAlaGlu---AspGlnMetLeuGluAr 673
 QY 603 CAGCGTGGAAAGCGCTTCTGAGACCAAGAGCGGCGCCAGTCTGTCTTCACAGCTGTGAA 662
 DB 673 gluValGlu-----LysGluAspGlnSerPheProArgSerProGluGluGluAspGI 691
 QY 663 GGAA-----CGACGAGCTCTTCTGAAAGGTGAAGGCGGATCCAGAGTCCGCTCGGAGAA 716
 DB 691 nGluAlaCysArgProLeuGlnLysGlu-----AsnGlnGluProLe 705
 QY 717 CTCCGATATGACAGCAGTGGCTGACCTTGAAAGAGCTTCAAGATCGCTTCGAGGAA 776
 DB 705 uGlyTyTrGluGluAlaGluGly-----GlnIleLeuGluArgLeuIleGluLy 721
 QY 777 GCGGAAACGAGGACCCACTGAGAGCGCTGAAAGGATCCAGAGTCCGCTCGGAGAA 836
 DB 721 sGluSerGlnGluSerLeuArgSerProGluGluGluAspGlnGluAlaGlyArgSerLe 741
 QY 837 GCGCGGAGGAGGCTCCGCGGAGCTGTGGC---TCCGAGGCGAGTGCACACTGTGGA 893
 DB 741 uGlnLysGluAsn-----GlnGluProLeuGlyTyTrGluGluAlaGluAspGlnMetLe 759
 QY 894 GGGCTCTCTGCCAGTAAGCAGGAGCGCGGAGCAACATCAGCGGGTTGTGCTCC----- 946
 DB 759 uGluArgLeuIleGluLysGluSerGlnGlu-SerLeuLysSer-----ProGluGlu 777
 QY 947 -----AGGCTGGGAAGATGACAGAGAGTAAAGTTGGAGGGAAGGCGG----- 991
 DB 777 snGlnArgIleGlyLysProLeuGluArgGluAsnGlnArgProLeu----- 813
 QY 992 -----CTCAGGACATCAAGATGAGGAGCGCTCGAGACTTGGG 1028
 DB 797 luAsnGlnGluThrPheValProLeuGluSerArgAsnGlnArgProLeu----- 813
 QY 1029 CCGACCGAGCCTGAATGTGAGGTTTACGCCCAACCGCTGATTGTCATTTCGCGCCA 1088
 DB 814 -----ArgSerLeuGluValGluGlu-GluGluGlnArgIleValLysProLeuGluLys 831
 QY 1089 GCCTCACACACAGGCTTTATGATTG-----CTGTGACCGCTGGAAGAATG 1136
 DB 832 ValSerGlnAspSerLeuGlySerLeuAlaGluGluAsnValGlnProLeuArgTyLeu 851
 QY 1137 GTTTCATGCGGATTGTGCGGCAATTCT-----GAGCTCGAGG 1175
 DB 852 Glu-GluAspAspCysIleAsnLysSerLeuLeuGluAspLysThrHisLysSerLeuGI 871

; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 22507
 ; LENGTH: 341
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-22507

Alignment Scores:
 Pred. No.: 3,11e-05 Length: 341
 Score: 162.00 Matches: 97
 Percent Similarity: 37.70% Conservative: 21
 Best Local Similarity: 30.99% Mismatches: 114
 Query Match: 3.45% Indels: 81
 DB: 4 Gaps: 19

US-09-787-016A-1 (1-2610) x US-09-252-991A-22507 (1-341)

QY	270	CGCAAGGCGACCGGAGCAATGAGGCGACCTAAGGCCATCAAAACCCACCGACAAAGA	329
DB	3	ArgGlnHisGlnProGlnGlu-----GlnProHisA-SerGlnArg	15
QY	330	GTT-----CAGGAAACATGGGTTTTTCGAAGGACCACTATCCCAAGCG	374
DB	16	ValTyrArgGlnAlaGluAspAlaArgSerGlyArgSerArgLeuHisGlnA-galaVal	35
QY	375	AGAGGCGCAGGCGACGCGAGGCTGACCCACTGGAGCGCCGCCACCCACGACGACGCT	434
DB	36	ArgProArgArgGlyArg-----	41
QY	435	GGGCTCTCCCTCGCGGCGAGTGGAGCGACGCCAAGCGCACTGAGCGCGTGGAGCAGTT	494
DB	42	--ProArgProAlaAlaGlnHisProGluGlyArgHisHis--GlyArgGlyAla--	58
QY	495	CTCAGCACTGCGGGCGCGCGGAGGAGCATGCTCTCTCCCTGGAGGATTCGTGG	554
DB	59	ProHisHis-----ArgGluProGluGlyLeuHisGlnGlyGlyGlnGly	73
QY	555	-----TGAGGCCACGCTCTGCCCGCGC---CACAGAGCGCCGAGAC---AGCCTC	595
DB	74	SerArgGlyValHisGlnHisArgLeuProArgProHisArgArgAspProHisLeu	93
QY	597	CGAGGCGAGCGTGAAGACGCTCTGAGACACAGAGCGGCCCGCAGCTCTCTCCACAGC	656
DB	94	HisGlySerArgArgArg-GlyAlaGlnGlyArgHisGluValGlyValAspArgAr	113
QY	657	TGTCAGAGAA-----CGACAGCGCTCTTCTGAAAGAGTGAAAGAGGGGA	701
DB	113	gLeuArgGluGlnGlnArgArgArgProGlyHisArgProAlaArgGlnGlyAlaAs	133
QY	702	TGACACGATGACACCTCCGATGATGACGCGATGCGCTGACCTTGAAGAGCTTCAGAA	761
DB	133	paArgGlnGlyHisValGlyHisAlaArgProAspGlyArgHis--AlaArgAlaGluAsp	152
QY	762	TCGCGCTTCGAGGAGCGGAAACAGGAGCCACTGAGAGGCCCTGAAGGGATCCAGAG	821
DB	153	ArgProSer-----AlaGlyArgArgGlnHisArgLeuGlyProValAlaAspArg	170
QY	822	TCGCGCTCGGAGAAAGCG-----CGGGAGGAGGGTCCCGCGAGACTGTGGGCTCCGAGGC	878
DB	171	AspProAlaArgProAlaLeuProGlnGlyArgArgValArgProGlyArgAlaGly	190
QY	879	CAGTGACAC-----TGTGAGGGCGGT---CTTGGCCAGTAAAGCAGAGCGC-----	920
DB	191	GltAlaHisProGlyLeuGlyArgHisProAspHisSerAlaGlyAlaGluHisGln	210
QY	921	-----CGAGAACCATCAGGGGGTGTCTCCAGGCTCGGGAAGATGA	962
DB	211	LeuAspArgGlyArgAspGlnGluArgGlyGlyGlnGlnArgProGlyHisProArgLeu	230
QY	963	CAGACAGAGTAAGTTGGAGGAAAGCG-----GGCTCAGGACATCAAAGA	1007

1176	Qy	GAGGCTTTTGGAAAGGAA	TGGGGAA	GACTATATCTGCCCAA	CTGCA	CACCATTTCTGCAAGT	1233
871	Db	yserLeuGluAspArgasnGlyaspserIleIlelleProGlnGluSerGluThrGlnIva	891				
1236	Qy	G-----CAGGATGAG-----	ACTCATTTCA	GAAACGC	GAGA	1265	
891	Db	IserLeuAvpProGluGluGluAspGlnArgIleValAsnHisLeuGluLysGluSe	911				
1266	Qy	TCAG-----CAGGAAGCTAAATGGAGACCTGGAGATGCTGATGG	1304				
911	Db	rGlnGluPheSerArgSerSerGluGluGluGluGlnValMetGluArgSerLeuGluGln	931				
1305	Qy	CACCGATTGTCAAGTATAGCAACAATAGACAGAGCTCTAGCGAAGACCAAGGGATAAA	1364				
931	Db	yGluAsnHisGluSerLeuSerSerValGluLys-----GluaspGlnMetValGln	948				
1365	Qy	GGGTAGAAATGGAAGCTCCAAATCCAACTGGCAGAGAAGAACTCAAGATCTTCAGACC	1424				
948	Db	userGlnLeuGluLysGluSerGlnAspSerGlyLysSerLeu-----	962				
1425	Qy	TGTTGATAGAGCGCCTGGTGCCTCAAAATGTTATTTGGCCCGGGTGCTGTACCTGGCGCA	1484				
963	Db	-----GluaspGluSerGlnGluThrPheGlyPro-----	972				
1485	Qy	GCCCGACTCGTGTACTGCAGTAAATGACTGTATCTCTCAAA	1544				
973	Db	-----LeuGluLysGluAsnAlaGluSerLeuAr	982				
1545	Qy	GTTTCTTAAGCTCAGGTAAAGAACAGACAGACCAAGACCTTAAGAAAGATGAAGTG-----	1599				
982	Db	gserLeuAlaGlyGlnAspGlnGluGlnLysLeuGluGlnGluThrGlnGlnThrLe	1602				
1600	Qy	-----AAGCCAGAGAAG-----CCCAGTCT	1619				
1002	Db	uArgAlaValGlyAsnGluGlnMetAlaValSerProGluLysValAspProGluLe	1022				
1620	Qy	TCGAAA-----TGGCGTCTCAGGCAGGTAT	1646				
1022	Db	uProLysProLeuGlyAsnAspGlnGluIleAlaArgSerLeuGlyLysGluAsnGlnGln	1042				
1647	Qy	TAAATCTCTTCTGTGCACAAGAGACAGCTCCAGAAAAAAGAGACACAGCTGAAGAA	1706				
1042	Db	userLeuValSerLeu-----LysGluLysGlyIleGluThrValLysSe	1057				
1707	Qy	G-----GCAGTGTGTCTCCCTCGCGGAGT-----GAAGACACTCGGAAGGAAGC	1751				
1057	Db	rLeuGluThrGluIleGluProLeuGluThrAlaGluGluAspLeuGluArgGly	1077				
1752	Qy	AGTTGTGACAGACAGCAGCGCTCGTGGCGAGCGATCACAATTCACATGCAGTAAAGCC	1811				
1077	Db	sSerIleAspThrGlnGluProLeuThrSerThrGluValAlaArgGluThrValGluPr	1097				
1812	Qy	AGAAAAGACGCTGCTCCCTCGCGCGTCACTGTTGTATAAATGTATGTATCACCTAGGGGT	1871				
1097	Db	oProGluAspGluProGly-----SerLe	1106				
1872	Qy	TGGCCTCTCGGACCCCTCCCGT	1893				
1106	Db	uGlySerValAspGluAsnArg	1113				

RESULT 9
 US-09-252-991A-22507
 ; Sequence 22507, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788

Db 7325 -----ThrProAlaProGluAlaProAlaGlu 7333
 QY 556 GAGCCACGCTGCGCCGCC----- 576
 Db 7334 GlnProLysProAlaProAlaProGlnProAlaProAlaProLysProAla 7353
 QY 577 -----ACAGCCGCGAGACAGCTCCGAGGCGAGCGTGGAAAGC 615
 Db 7354 GluGlnProLysAlaGluLysThrAspArgGlnGlnAlaGluGluAspTyrAlaArg 7373
 QY 616 GCTCTCTGAG-----ACCAGAGCGCGCCCGCTCTGCTTCCACAGCTGTG 660
 Db 7374 SerGluGluGluTyrAsnArgLeuThrGlnGlnProProLysAlaGluLysProAla 7393
 QY 661 -----AAGAAACACACGCTCTTCTGAAAAGCTGAAGAGGAGGATGACCAC 708
 Db 7394 ProAlaProLysProGluGlnProAlaProAlaProLysAsnSerLysGlyGluGlnAla 7413
 QY 709 GATGACACCTCCGATAGTCACAGCGATGCTGCTTCAAGAGAGCTTCAGAAATCGCCTT 768
 Db 7414 GluGlnTyrArgSerAlaAlaGlyGlyAspLeuAlaAlaLysGlnValGluLeuGluLys 7433
 QY 769 CCG-----AGGAAGCGGGAACAGAGCCCGCTGAGAGCGCCCTCTGAAGGGATC 816
 Db 7434 ThrGluAlaAspLeuLysLysAlaValAsnGluPro---GluLysProAlaProAla--p 7452
 QY 817 CAGAGTCGCTCGGAGAGCGCGGAGGAGGCTCCCGCGAGACTGTGGGCTCCGAG 876
 Db 7452 xGluThrProAlaProGluAlaProAla--GluGlnProLysProAlaProAlaProGln 7471
 QY 877 GCACGTGACACTGTGAGGCGCTCTGCGGAGGAGGAGCTTGGGCGCGAGCGCTGAA 1044
 Db 7511 ThrGlnGlnProProLysAlaGluLysProAlaPro-----AlaProGlnProGlu 7528
 QY 1045 TGTAGGGTTACAGCCCGGCGCTGTATTGCTTGGCTTGGCGCGAGCTTCAACAACAGG 1104
 Db 7529 GlnProAlaProAlaProLysSerLeuLysGluLeuAspGluSerAspSerGlu----- 7546
 QY 1105 TTTATGATTGCTGACCGCTGAGAGATGTTTCATGCGGATTTGTGGGCAATTCT 1164
 Db 7547 -----AspTyrValLysGluLysLeuGluGluSerAspLysLys----- 7561
 QY 1165 GAGGCTCGAGGAGGCTTTTGGAAAGGATGGGAA-----GACTATATCTGCCCAAC 1218
 Db 7562 AspAlaLysGlnAlaLysLeuSerLysLeuGluGluSerAspLysLys----- 7578
 QY 1219 TGCACCATTTGCAAGTCAGGAGTACCTATTCAGAACCGCA-----GAT 1266
 Db 7579 -----AspGluLeuAspAlaGluLeuAlaLysLeuGluLysAsp 7591
 QY 1267 CAGCAGGAAGCTAAATGAGACCTGAGATGCTGATGGCAGGATTTGTACAGTATAGGA 1326
 Db 7592 ValGluAspPheLys---SerAspGlyGluGlnAlaGly-----GlnTyrLeuAla 7608
 QY 1327 ACAATAGCAG-----AAGTCTAGGAGACACCAAGGATAAGGCTAGATT 1374
 Db 7609 AlaAlaGluGluAspLeuLeuAlaLysLysLeuGluGluGlnThrGluAlaAspLeu 7628
 QY 1375 GAGAAAGCTCAATCCAAAGTGGCGAAGAACTCAAGATCTTC-----CAGCCTGTG 1428
 Db 7629 LysLysAlaValAsnGluProGlyLysProAlaProAlaProAlaProGluThrProAla 7648
 QY 1429 ATAGAGCGCTGCTGCTCAAAATGATTGGGCCCGGCTGTGCTACGTGGCG----- 1482
 Db 7649 ProGluAlaProAlaGluGlnProLysProAlaProGluThrProAlaProAlaProLys 7668

QY 1483 -----CAGCCGACTCGTCTACTGCACTAAT 1509
 Db 7669 ProGluLysProAlaGluGlnProLysProGluLysProAlaAspGlnGlnAlaGluGlu 7688
 QY 1510 GACTGTATCTCAAAACAGCGCGAGACATGAAGTTTCTAAGCTCAGGTAAGAAGACAG 1569
 Db 7689 Asp-----TyrAlaArgSerGluGluGluTyrAsnArgLeuThrGlnGln 7704
 QY 1570 AAGCAAGAGCTAAAGAAAGATGAGATG-----AAGCCAGAGAGAGCCAGCTTCCG 1623
 Db 7705 GlnProAlaProAlaGlnLysProGluGlnProAlaLysProGluLysProAlaGluGlu 7724
 QY 1624 AAATCGGCTGCTCAGCGCAGGATTAATAATCTCTCTGTGCAC----- 1665
 Db 7725 ProThrGlnProGluLysAspAlaGluLeuAlaLysLeuGluLysAsnValGluTyrPhe 7744
 QY 1665 ----- 1665
 Db 7745 LysLysThrAspAlaGluGlnThrGluGlnTyrLeuAlaAlaGluLysAspLeuAla 7764
 QY 1666 ---AAGAGACAGCTCCAGAAAAAAGAGACACAGCTGAGAGAGAGCTGTGTCTCCT 1722
 Db 7765 AspLysLysAlaGluLeuGluLysThrGluAlaAspLeuLysLysAlaValAsnGluPro 7784
 QY 1723 GCGCGAGTGAAGCACTCGGAGAGAGAGAGCTTGTGAGAGAGAGAGAGAGAGAGAG 1782
 Db 7785 -----GluLysProAlaGluGluThrProAlaProLysProGluGlnPro 7801
 QY 1783 AGCGATCACAAATCAATGACAGTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1836
 Db 7802 AlaGluGlnProLysProAlaProAlaProGlnProAlaProAlaProLysPro 7819
 RESULT 12
 US-10-164-595-56
 ; Sequence 56, Application US/10164595
 ; Patent No. 6657054
 ; GENERAL INFORMATION:
 ; APPLICANT: Origene Technologies, Inc
 ; TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
 ; FILE REFERENCE: 1U 103 R1
 ; CURRENT APPLICATION NUMBER: US/10/164,595
 ; CURRENT FILING DATE: 2002-06-10
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 56
 ; LENGTH: 779
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-164-595-56
 Alignment Scores:
 Pred. No.: 8,98e-05 Length: 779
 Score: 158.50 Matches: 106
 Percent Similarity: 32.67% Conservative: 59
 Best Local Similarity: 20.99% Mismatches: 195
 Query Match: 3.38% Indels: 145
 DB: 19
 US-09-787-016A-1 (1-2610) x US-10-164-595-56 (1-779)
 QY 268 GAGCAAAAGCGCGAGCCGAGCAATGAGGAGGAGCCATCAAGCCATCAAAACCCAGCAAA 327
 Db 271 GluAspLysArgAspLeuLysSerArgGluLysSerLys----- 283
 QY 328 GAGTTCAAGAAACATGGGGTTTTCGAGGAGACCATATCGCAAGCGAGAGGCGCAGGG 387
 Db 284 -----PheArgAspThrHisLysLysLeuGluGluLysGly 296
 QY 388 GAGCGGAGCTGACCCACTGAGCGCGCCACCCACAGCAGCAGCTGGGCTGTCTCCTG 447
 Db 297 LysLysGluLysGlu----- 301

2086	QY	GGTGAACGCTTACCTTAAATGATTGATCTTTAATCTCTGTTTTCTCTCTCAGGCTCTGGT	2145
719	Db	urhAlaileHis-----CysphenisGluAlaargLeuAs	731
2146	QY	AGATATTGTGATTTCTCTTATCCAGTCTGATTGATGACCACTGTCGGGACAGCCAC	2205
731	Db	pAspGluCysAlaIaPheTyr-----ThrSerAlaIaSerPr	743
2206	QY	ATCCACCCCT-----GTCTGCACATGAGTTGTTCTGACA	2239
743	Db	oSerGlyProThrArgValCysThrAsnProValAlaThr	756

RESULT 14

US-08-751-965-5

US-08-151-203-3
: Sequence 5, Application US/08751965

Sequence 3, Appendix
Patent No. 5858360

FACEBOOK NO. 3836360
GENERAL INFORMATION:

APPLICANT: Fukuda, Michiko N.

APPLICANT: FUKUDA, MICHIO N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting

	TITLE OF INVENTION:	TITLE OF INVENTION:	TITLE OF INVENTION:
1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			
16.			
17.			
18.			
19.			
20.			
21.			
22.			
23.			
24.			
25.			
26.			
27.			
28.			
29.			
30.			
31.			
32.			
33.			
34.			
35.			
36.			
37.			
38.			
39.			
40.			
41.			
42.			
43.			
44.			
45.			
46.			
47.			
48.			
49.			
50.			
51.			
52.			
53.			
54.			
55.			
56.			
57.			
58.			
59.			
60.			
61.			
62.			
63.			
64.			
65.			
66.			
67.			
68.			
69.			
70.			
71.			
72.			
73.			
74.			
75.			
76.			
77.			
78.			
79.			
80.			
81.			
82.			
83.			
84.			
85.			
86.			
87.			
88.			
89.			
90.			
91.			
92.			
93.			
94.			
95.			
96.			
97.			
98.			
99.			
100.			

FILE OF INVENTION:	FILE
NUMBER OF SEQUENCES:	22

NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

CITY: San Diego
STATE: CaliforniaSTATE: CALIFORNIA
COUNTRY: USACOUNTRY: 01
ZIP: 92122

ZIP: 52122
COMPUTER READABLE FORM:

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC compatible

COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS

```

;
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0.  Version #1.25

```

```

; SOFTWARE: PALCHIN REL
; CURRENT APPLICATION DATA:

```

7 CURRENT APPLICATION DATA: US/08/751.965

APPLICANT NUMBER: 0
ESTLING DATE: Herewith

FILE NO. 100-441101-100
FILING DATE: 11/11/2011
CLASSIFICATION:

CLASSIFICATION: UNCLASSIFIED
 REPORT APPLICATION DATA: UNCLASSIFIED

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439 818

APPLICATION NUMBER: US 01
FILING DATE: 13-MAY-1995

FILED DATE: 12-MAY-1995
ATTORNEY / AGENT INFORMATION:

NAME: Campbell Cathryn A

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31 81

REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-

TELECOMMUNICATION INFORMATION

TELEPHONE: (619) 535-9000

TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 5

SEQUENCE CHARACTERISTICS:

; LENGTH: 778 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

MOLECULE T

Alignment Scores:	
Pred. No.:	0.000132
Score:	156.50
Percent Similarity:	30.53%
Best Local Similarity:	21.36%
Query Match:	3.34%
DB:	2
Length:	778
Matches:	170
Conservative:	73
Mismatches:	272
Indels:	282
Gaps:	38

US-09-787-016A-3 (1-2610) x US-08-751-965-5 (1-778)

Qy	302	CTAAGGCATCA-----AACCCACGCGCAAGAGTTCAGGA	337
Db	95	LeuArgProSerProArgGlyGlyLnaSnValGlyProGlyProProAlaGlnThrGluAla	114
Qy	338	AAACATGGGGTTTCGAGGACCACTATCGCCAGCGAGGGCG-----	362
Db	115	Pro---GlyThrIleGluPheValAlaAspProAlaAlaLeuAlaThrIleLeuSerGly	133

QY 1235 TGCAGGATGAGACTCATTCAAAACGCGAGATCAGCAGGAGAGCTAAATGAGACCTGGAG 1294
DB 457 - - - - - LeuAsnGlyGlySer 462
QY 1295 ATGCTGATGCCACCGATTGTACAGTATAGACAAATAGACGAGAGTCTAGCGAAGACC 1354
DB 463 LeuAspMet - - - - - 465
QY 1355 AAGGGATAAGGTTAGATTGAGAAAGCTGCAAAATCCAAAGTGGCAAGA - - - - - 1402
DB 466 - - - - - ValGluLeuGlnProLeuLeuThrGluLeuSerArgThrLeuAsnAla 481
QY 1403 - - - - - AGAACTCAAGATCTTCAGCGCTGTG - - - - - ATAGAGCGCGCTGGTCTCTCA 1449
DB 482 ThrGluHisAsnSerGlyThrSerHisLeuProGlyLeuLeuHisSerGlyLeuPro 501
QY 1450 AAA - - - - - TGTATTGGCCCGGGTGTCTCAGCTGGCGCAG - - - - - 1485
DB 502 LysProCysLeuProGluGluCysGlyGluProGlnProCysProAlaGluProGly 521
QY 1486 CCGGACTCGGTGACTGACGTAATGACTGATCTCAACACGCGCAGCGACAATGAG 1545
DB 522 ProProGluAlaPheCysArgSerGluProGluLeuProGluProSerLeuGlnGluGln 541
QY 1546 TTCTAAGCTCAGTAAAGAACAGAACGCAAGCCTAAAGAAAGATGAAGTGAAGCCA 1605
DB 542 Leu - - - - - GluValProGluProTyProProAlaGluProArgPro 555
QY 1606 - - - - - GAGAACCCAGTCTTCGAAATGCGGTCTCAGGAGGATTT 1647
DB 556 LeuGluSerCysCysArgSerGluProGluLeuProGluSerArgGlnGluGlnLeu 575
QY 1648 AAAATCTCTCTGTGCACAGAGACGACCTCCAGAAAAAAGAGACCAAGTGAAGAG 1707
DB 576 GluVal - - - - - ProGluProCysPro - - - - - 582
QY 1708 GCAGTGTGTCTCTGCGCGAGTGAAGCACTCGGGAAGAACGAGCTTGTGAGAGCAGC 1767
DB 583 - - - - - ProAlaGluProArgProLeuGluSerTyCysArgIleGluProGlu 598
QY 1768 ACCCGTCTGTTGGGAGCGATCAATTAATGACGTAAGCAGCAAAAGATCTGCT 1827
DB 599 IleProGluSerArgGlnGlnGlnLeuGluValProGluProCysProAlaGlu 618
QY 1828 CCTCGCGCTCACTGTTGTATAATGATGTATCACCCTAGGGGTGGCTCTCTGGACCCC 1887
DB 619 ProGlyPro - - - - - LeuGlnPro 624
QY 1888 TCCGTTCTTCTGATPAGCATCCCTGGGCGCTGTCCA - - - - - GGACTGGGAGTTGCA 1941
DB 625 SerThrGlnGlnSerGlyProProGlyProCysProArgValGluLeuGlyAlaSer 644
QY 1942 GCTTTGTTAAGCTGATCAGAG - - - - - 1965
DB 645 GluProCysThrLeuGluHisArgSerLeuGluSerSerLeuProProCysCysSerGln 664
QY 1966 - - - - - CACCGGCTGCACCATCAGCGGGAGCAGACGCC 1998
DB 665 TrpAlaProAlaThrThrSerLeuLeuPheSerSerGlnHisProLeuCysAlaSerPro 684
QY 1999 - - - - - ATGTCAGATGCTCTCTCTGCTGCC - - - - - CTG 2025
DB 685 ProIleCysSerLeuGlnSerLeuArgProProAlaGlyGlnAlaGlyLeuSerAsnLeu 704
QY 2026 TGTCCATCCCTAGTCTCTCAGGACTCTCTGCTCACTGTTTTCCAAAGCTGTAAACCTCACT 2085
DB 705 AlaProArgThr - - - - - LeuAlaLeuArgGlu-SerLeuLysSerCysLe 719
QY 2086 GGTGAAGCTTCACTTAATGATTGATTCTTAACTCTGTTTCACTCTCAGGCTCTGCT 2145
DB 719 uThrAlaIleHis - - - - - CysPheHisGluAlaArgLeuAs 731
QY 2146 AAGTATTGTATTCTTCTTCTTCCAGTCTGATTGATGATGATGATGATGATGATGATGAT 2205

DB 731 pAspGluCysAlaPheTy - - - - - ThrSerArgAlaSerPr 743
QY 2206 ATCCACCCCT - - - - - GTCTGCACATGAGTTGTTCTGACA 2239
DB 743 oSerGlyProThrArgValCysThrAsnProValAlaThr 756
RESULT 15
US-08-738-975-5
; Sequence 5, Application US/08738975
; Patent No. 5880267
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
; TITLE OF INVENTION: Trophinin
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/738,975
; FILING DATE: herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/439,818
; FILING DATE: 05-Dec-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 2251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 778 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-738-975-5
Alignment Scores:
Pred. No.: 0.000132 Length: 778
Score: 156.50 Matches: 170
Percent Similarity: 30.53% Conservative: 73
Best Local Similarity: 21.36% Mismatches: 272
Query Match: 3.34% Gaps: 38
DB: 2
US-09-787-016A-1 (1-2610) x US-08-738-975-5 (1-778)
QY 302 CTAAGGCCATCA - - - - - AACCCACGACCAAGAGTTCCAGGA 337
DB 95 LeuArgProSerProArgGlyGlnSerValGlyProGlyProAlaGlnThrGluAla 114
QY 338 AAACATGGGGTTTTCGAGGACCACTATCGCAAGCGAGAGGGCG - - - - - 382
DB 115 Pro---GlyThrIleGluPheValAlaAspProAlaAlaLeuAlaThrIleLeuSerGly 133
QY 383 CAGGGACCGCGGAGGCTGACCCACTGGAGCCGCCACCGCCAGCAGCTGGGCTGT 442
DB 134 GluGlyValLysSer---CysHisLeuGlyArgGlnProSerLeuAlaLysArgValLeu 152
QY 443 CCTGTGGGCGCAGTGGGAGGAGC - - - - - 466

Search completed: April 28, 2004, 11:03:03
Job time : 89.5 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2989	63.7	562	3	AAy67579	Human dea
2	2981	63.6	562	4	AAy93638	Human pro
3	2432	51.9	1191	4	AAm80219	Human pro
4	2352.5	50.2	775	4	ABg22389	Human pro
5	2311	49.3	647	4	AAm79235	Human pro
6	2154.5	45.9	514	3	AAy67580	Murine de
7	1551	33.1	330	4	ABg22388	Human hum
8	958	20.4	181	3	AAb43724	Human can
9	898	19.2	167	4	ABg22387	Human hum
10	494	10.5	108	5	ABp05496	Human ORF

PR 17-SEP-1998; 98US-01008/3P-

XX (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
 PA (BANN/) BANNERMAN D G.
 XX Alonso CM, Domingo DG, Grandien A, Leonardo B, Martinez P;
 PI WPI; 2000-271426/23.
 DR N-PSDB; AA290578.
 XX New DNA encoding human and murine death inducer-obliterator 1
 PT polypeptides, useful in the treatment of cancer, autoimmune diseases,
 PT diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
 PT hyperproliferative skin disorders.
 XX Claim 6; Fig 1C; 27pp; English.
 XX The invention provides nucleic acids encoding the human and murine death
 CC inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be
 CC expressed by standard recombinant methodology. The DIO-1 polypeptides,
 CC agonists and antagonists are used as a medicament for treating diseases
 CC characterized by an alteration in cell death or by hyperproliferation,
 CC e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
 CC tumours, malignant tumours or hyperproliferative skin disorders. They are
 CC also useful in the treatment of metabolic, proliferative or inflammatory
 CC conditions. The present sequence represents the human DIO-1 polypeptide
 XX SQ Sequence 562 AA;

Alignment Scores:
 Pred. No.: 2,82e-224 Length: 562
 Score: 2989.00 Matches: 562
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.74% Indels: 0
 DB: 3 Gaps: 0

US-0787-016A-1 (1-2610) x AAY67579 (1-562)

QY 265 ATGAGAGCAAGGCGGACCGGACCAATGAGGAGCGCACTAAGCCATCAAAACCCACGAG 324
 DB 1 MeAspAspLysGlyAspProSerAsnGluGluAlaProLysAlaLeuLysProThrSer 20
 QY 325 AARAGAGTTCAGGAAACATGGGGTTTTTCGAGGAGCACCACTATCCCAAGCGAGAGGCGCA 384
 DB 21 LysGluPheArgLysThrTrpGlyPheArgArgThrThreAlaLysArgGluGlyAla 40
 QY 385 GGGAGCGGAGCGTACCCACCTCGAGCGCCGACCCGCCACGACGAGCTGGCGCTGTCC 444
 DB 41 GlyAspAlaGluAlaAspProLeuGluProProProGlnGlnGlnLeuGlyLeuSer 60
 QY 445 CTGCGGCGAGTGGGAGCGAGCCCAAGCGCACTGAGCGCGTGGAGCAGTTCCCTGACCAT 504
 DB 61 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGlnPheLeuThre 80
 QY 505 GCGGCGCGCGCGGAGGAGAGTCCCTGTCTCCCTGAGAGATTCGTGAGCCGAG 564
 DB 81 AlaArgArgGlyArgArgSerMetProValSerLeuGluAspSerGlyGluProThr 100
 QY 565 TCCTGCCCCCGCACAGACGCGGAGACAGCTCCGAGGCGGAGCGTGGAAAGCGCTTCGAG 624
 DB 101 SerCysProAlaThrAspAlaGluThrAlaSerGluGlySerValGluSerAlaSerGlu 120
 QY 625 ACCAGAGCGGCGGCGGAGTACCGAGTACAGCTCCGATAGTACAGCGATGCGCTTCGAA 684
 DB 121 ThrArgSerGlyProGlnSerAlaSerThrAlaValLysGluArgProAlaSerSerGlu 140
 QY 685 AAGGTGAAGGCGGAGTACCGAGTACAGCTCCGATAGTACAGCGATGCGCTTCGAG 744
 DB 141 LysValLysGlyLysAspPheHisAspPheHisAspPheHisAspPheHisAspPheHis 160
 QY 745 TTGAAGAGGCTTCAGATCCGCTTCGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 804
 DB 161 LeuLysGluLeuGlnAsnArgLeuArgArgLysArgGluGlnGluProThrGluArgPro 180

QY 805 CTGAAAGGGATCCAGAGTCGCTCCGGAAGAAGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 864
 DB 181 LeuLysGlyIleGlnSerArgLeuArgLysLysArgGluGluGluGluGluGluGluGlu 200
 QY 865 GTGGGCTCCGAGGCGGAGTACAGCTGTGGAGGCGCTCTGCCAGTAAAGACAGAGCCGAG 924
 DB 201 ValGlySerGluAlaSerAspThrValGluGlyValLeuProSerLysGlnGluProGlu 220
 QY 925 AACCATCAGGGGGTGTGTCTCCAGGCTGGGAAAGATGACAGAGAGTAAAGTTGGAGGA 984
 DB 221 AsnAspGlnGlyValValSerGlnAlaGlyLysAspAspArgGluSerLysLeuGluGly 240
 QY 985 AAGCGGCTCAGGACATCAAAATGAGGAGGCTGGAGCTTGGGCGGAGCCGAGGCTCAA 1044
 DB 241 LysAlaAlaGlnAspIleLysAspGlnGluProGlyAspLeuGlyArgProLysProGlu 260
 QY 1045 TGTGAGGTTACGACCCCAAGCGCTGTATTTCATTTGCCGCCAGCTCCACAAACAGG 1104
 DB 261 CysGluGlyTyAspProAsnAlaLeuTyCysIleCysArgGlnProHisAsnAsnArg 280
 QY 1105 TTTATGATTCGTGACCGCTGTGACAAAGTGGTTTCATGGCGATTGTTGGGCTTTCT 1164
 DB 281 PheMetIleCysCysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSer 300
 QY 1165 GAGGCTCGAGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCCAACTGCACC 1224
 DB 301 GluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyIleCysProAsnCysThr 320
 QY 1225 ATTCTGCAAGTGCAGGATGAGACTTCATTGAGAAAGCGGAGATCAGCAGAGGAGTAAAGG 1284
 DB 321 IleLeuGlnValGlnAspGluThrHisSerGluThrAlaAspGlnGlnGluAlaLysTrp 340
 QY 1285 AGACCTCGAGATGCTGATGGCACCGATTGTCAAGATATAGAAACAAATAGAGCAGAGTCT 1344
 DB 341 ArgProGlyAspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSer 360
 QY 1345 AGCGAAGCAAGGAGTAAAGGTAGAAATTGAGAAAGCTGCAAATTCGAAGTGCAGAGAG 1404
 DB 361 SerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLys 380
 QY 1405 AAACCTCAGAGATTCACGCTGTGTATAGAGCGCTGTGTCTCAAAATGTTATTTGGCCCC 1464
 DB 381 LysLeuLysIlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyPro 400
 QY 1465 GGGTCTGTCTCAGTGGCGCAGCCGAGCTCGGTGTACTCCAGTAATGACTGTATCTCAAA 1524
 DB 401 GlyCysCysHisValAlaGlnProAspSerValTyCysSerAsnAspCysIleLeuLys 420
 QY 1525 CAGCGCGGAGGAGCAATGAGCTTTTCTAGCTCAGGTAAAGAACAGACAGCAAGCCTAA 1584
 DB 421 HisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys 440
 QY 1585 GAAAAGATGAAGATGAAGCAGAGAGCCGAGCTTCCGAAATGCGGTCTCAGGAGGT 1644
 DB 441 GluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGly 460
 QY 1645 ATTAAATCTCTCTGTGTCAGAGACAGCTCCAGAAAAGAAAAGAGACACAGCAGTGA 1704
 DB 461 IleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThrValLys 480
 QY 1705 AAGGAGGTGGTGGTCCCTCGCGGAGTGAAGCAGCTCGGGAAGGAGGAGCAGCTTGTGAG 1764
 DB 481 LysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaCysGluSer 500
 QY 1765 AGCAGCGGCTGTGGGAGGAGGATCAATTCATTCATGATGATGATGATGATGATGATGAT 1824
 DB 501 SerThrProSerTrpAlaSerAspHisAsnTyAsnAlaValLysProGluLysThrAla 520
 QY 1825 GCTCCCTCGCTCTGTTGTATTAATGATGATGATGATGATGATGATGATGATGATGATG 1884
 DB 521 AlaProSerProSerLeuLeuTyLysCysMetTyHisLeuGlyValGlyLeuLeuAsp 540

QY	1885	CCTCCGCTCTTCTGGATAGCATCCCTGGCCCTGTCACGACACGGAGTTGCAGCT	194
Db	541	ProSerArgSerPheTrpIleAlaIleProTrpAlaCysProGlyLeuGlyValAlaAla	560
QY	1945	TTGTGT	1950
Db	561	LeuCys	562
RESULT 2			
AAB93638			
ID	1	AAB93638 standard; protein; 562 AA.	
XX	AC	AAB93638;	
XX	AC		
DT	26-JUN-2001	(first entry)	
DE		Human protein sequence SEQ ID NO:13130.	
XX	Human;	primer; detection; diagnosis; antisense therapy; gene therapy.	
XX	Homo sapiens.		
PN	EP1074617-A2.		
XX	07-FEB-2001.		
PD			
XX	28-JUL-2000;	2000EP-00116126.	
XX	29-JUL-1999;	99JP-00248036.	
PR	27-AUG-1999;	99JP-00300253.	
PR	11-JAN-2000;	2000JP-00118776.	
PR	02-MAY-2000;	2000JP-00183767.	
PR	09-JUN-2000;	2000JP-00241899.	
XX			
XX	(HELI-)	HELIX RES INST.	
PA	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
PI	WPI;	2001-318749/34.	
DR			
XX			
XX	Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.		
PT	Claim 8;	SEQ ID NO 13130; 2537pp + Sequence Listing; English.	
XX			
CC	The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises:		
CC	(a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to the polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention		
CC	Sequence 562 AA;		
SQ			

Db 109 rAspSerAspSerAspGlyLeuThrLeuLysLeuGlnAsnAsgLeuArgArgLysAr 129
QY 780 GGAACAGAGAGCCACTGAGAGCCCTGAAAGGATCCAGATCCCTGCGGAAGAGCG 839
Db 129 gGluGlnGluProThrGluArgProLeuLysGlyIleGlnSerArgLeuArgLysLeAr 149
QY 840 CGGAGAGAGGCTCCGCGAGACTGTGGCTCCGAGGCCAGTGCACACTGTGAGGGCGT 899
Db 149 gArgGluGluGlyProAlaGluThrValGlySerGluAlaSerAspThrValGluGlyVa 169
QY 900 CTTGCCCTAGAGAGAGCCGAGAACCATCAGGGGTGTGTCCAGGCTGGGAAGA 959
Db 169 LeuProSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAlaGlyLysAs 189
QY 960 TGACAGAGAGTAAATTGAGGGAAGCGCTCAGGCATCAAGATCAAGTGAAGCCTGG 1019
Db 189 pAspArgGluSerLysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProGl 209
QY 1020 AGACTTGGCCCAACGAGCTGAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCAT 1079
Db 209 yAspLeuGlyArgProLysProGluCysGluGlyTyraAspProAsnAlaLeuTy-CysIl 229
QY 1080 TTGCGCCAGCTCAACAACAGAGTTATGATTTGCTGTGACCGCTGTGAAGATGCTT 1139
Db 229 eCysArgGlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluTrpPh 249
QY 1140 TCATCGCATTCGTGGGCTTTCTGAGGCTCAGGGAGCTTTTGGAAAGGAATGGGA 1199
Db 249 eHisGlyAspCysValGlyIleSerGluAlaArgGlyAsgLeuLeuGluArgAsnGlyGl 269
QY 1200 AGACTATATCTCCCAACTCACCATTCTGCAAGTGCAGAGTGAAGACTCATTCAGAAC 1259
Db 269 uAspTyrlleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluTh 289
QY 1260 GCAGATCAGCAGGAGCTAAATGAGACTCTGAGATGCTGATGCCACGATTGTACAG 1319
Db 289 rAlaAspGlnGlnGluAlaLysTrpArgProGlyAspAlaAspGlyThrAspCysThrSe 309
QY 1320 TATAGCAACAATAGACAGCAAGCTCTAGCAAGACCAAGGGATAAGGGTAGAATTGAGAA 1379
Db 309 rIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGluLyl 329
QY 1380 AGCTGCAAAATCCAGTGGCAAGAGAACTCAGATCTCCAGCTT----- 1425
Db 329 sAlaAlaAsnProSerGlyLysLysLysLeuLysIlePheGlnProGlyProGlyProVa 349
QY 1425 ----- 1425
Db 349 lProThrGlnLeuProValLeuTrpGlnValLeuGluIleAlaValSerArgSerIleSe 369
QY 1426 -----GTGATAGAGCGCCCTGGTGGCTTCAAA 1451
Db 369 rAlaPheThrLeuLeuHisCysIleSerCysLysValIleGluAlaProGlyAlaSerLy 389
QY 1452 ATCTATTGCCCGGCTGTCTCAGCTGGCGAGCCGACTCGTGGTACTGCAAGTAAATCA 1511
Db 389 sCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTy-CysSerAsnAs 409
QY 1512 CTGTATCTCTCAACACCGCCGACGCAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAA 1571
Db 409 pCysIleLeuLysHisAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnly 429
QY 1572 GCCAAGCCTTAAGAAGATGAGATGAGCCAGCAGAGAGCCAGCTTCCTCCGAATTCGCG 1631
Db 429 sProLysProLysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGl 449
QY 1632 TGCTCAGGAGGATTAAATCTCTTCTGTGCAAGAGACCAAGCTCCAGAAAAAAGA 1691
Db 449 yAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGl 469
QY 1692 GACCACAGTGAAGAGGACTGTGTCTCTCGCGGAGTGAAGCACTCGGAGGAGGC 1751

469 uThrThrValLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAl 489
1752 AGCTTGTGAGAGCAGCAGCGCTGCTGGCGAGCATCACAAATTACAAATGCAATAAGCC 1811
489 aAlaCysGluSerThrProSerTrpAlaSerAspHisAsnTyraAlaValLysPr 509
1812 AGAAAAAGACTCTCTCCCTCCCTCCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 1851
509 oGluLysThrAlaAlaProSerProSerLeuLeuTy-Tlys 522
RESULT 4
ABG22389 standard; protein; 775 AA.
ID ABG22389
XX AC ABG22389;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #22380.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX N-PSDB; AAS86576.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forsenics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 20; SEQ ID NO 52748; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forsenics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 775 AA;
XX Alignment Scores:

Prod. No.:	1,46e-174	Length:	775
Score:	2352.50	Matches:	485
Percent Similarity:	74.42%	Conservative:	24
Best Local Similarity:	70.91%	Mismatches:	58
Query Match:	50.17%	Indels:	117
DB:	4	Gaps:	14

us-09-787-016a-1 (1-2610) x ABG22389 (1-775)	
QY 202 GGGTTTTGGTGTGTTTATTTAGGTTTTCAGGAAAGTGTCCAAAGCTTTCAGTCTTCGAGCA 261	QY 1072 TATTGCATTTGCCGCCAGCCCTCAACAACAGAGTTTATGATTTTCTGTGACCGCTGTGAA 1131
DB 80 GlyGlyCysValAlaGlyLeuIleSerGlyLysSerValGlnAlaPheSerValGlyAla 99	DB 400 TyrCysIleCysArgGlnProHisAsnArgPheMetIleCysCysAspArgCysGlu 419
QY 261 ----- 261	QY 1132 GAATGGTTTCATGCGCATTTGTGGCATTTCTGAGGCTCGAGGAGCTTTTGGAAAGG 1191
DB 100 GlyArgTrpProLeuLeuArgValPheGlyGlyAsnValThrPheCysSerValValSer 119	DB 420 GluTrpPheHisGlyAspCysValGlyIleSerGlnAlaArgGlyAlaGluLeuGluArg 439
QY 262 ----- GGTATGGACGACAAAGCGCACCGAGCAAT 291	QY 1192 AATGGGAAGACTATATCTGCCCAAACTGCACCACTTCTGCAAGTGCAGGATGAGACTCAT 1251
DB 120 TrpSerLeuCysAlaTyrThrLeuProLeuGlyMetAspAspLysGlyAspProSerAsn 139	DB 440 AsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHis 459
QY 292 GAGGAGCACCTAAGCCCATCAAAACCACGACAAAGAGTTTCAGGAAACATCGGGTTT 351	QY 1252 TCAGAAACGGCAGATCAGCAGGAAGCTTAATGGAGACCT---GGAGATGCTGATGGCACC 1308
DB 140 GluGluAlaProLysAlaIleLysProThrSerLysGluPheArgLysThrTrpGlyPhe 159	DB 460 SerGluThrAlaAspGlnGlnGluAla**MetGlyAspLeuGlyAspAlaAspGlyThr 479
QY 352 CGAAGGACCACTATCCCAAGCGAGAGCGCGAGGACCGCGAGGCTGACCCACTGGAG 411	QY 1309 GATTGTACAAGTATAGGAACAATAGACGAGAAG---TCTAGCGAAGACCAAGGGATA 1362
DB 160 ArgArgThrThrIleAlaLysArgGluGlyAlaGlyAspAlaGluAlaAspProLeuGlu 179	DB 480 AspPheThrSerIleGlyAsnIleArgSerArgSerLeuSerArgProLysGlyLeu 499
QY 412 CGCCACCCACGACGAGCTGGGCGCTCTCCCGCGAGTGGGAGGAGCCCAAG 471	QY 1363 AAGGTAGAATTGAGAAAGCTGCAAAATCCAAAGTGGCAAGAAAGAACTCAAGATCTTCAG 1422
DB 180 ProProProGlnGlnGlnGlyLeuSerLeuArgSerGlyArgGlnProLys 199	DB 500 LysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysGluLeuLysIlePheGln 519
QY 472 CGCACTGAGCGCTGGAGCATTTCTTCACCATTCGCGCGCGCGCGAGGAGGAGCATG 531	QY 1423 CCT----- 1425
DB 200 ArgThrGluArgValGluGlnPheLeuThrIleAlaArgArgGlyArgSerMet 219	DB 520 ProValLeuGluIleAlaValSerArgSerIleSerAlaPheThrLeuLeuHisCysIle 539
QY 532 CTGTCTCTCGTGAAGATTCTGTGAGCCACGCTCTGCGCCGACACAGACCCGAGCA 591	QY 1426 -----GTGATAGAGCGCGCTGTGCTCTCAAAATGATATGCGCCCGGTGTGTCTAC 1476
DB 220 ProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaGluThr 239	DB 540 SerCysLysValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCysHis 559
QY 592 GCCTCCGAGGCGAGCTGGAAGCGTCTTGACACCAAGAGCGCGCCCGAGTCTGCTTCC 651	QY 1477 GTGGCGCAGCCGCTCGTGTCTGCGAATGCTGCTATCTCTCAACACGCGCCAGCG 1536
DB 240 AlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGlnSerAlaSer 259	DB 560 ValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLysHisAlaAlaAla 579
QY 652 ACAGCTGTGAAGAAACACAGCAGCTCTTCTGAAAGGTGAAGAGGAGGATGACCAAGAT 711	QY 1537 ACAATGAAGTTTCTAAGCTCAGGTAAGAAACAGAAAGCCAAAGCTTAAAGAAAGATGAAG 1596
DB 260 ThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyAspAspHisAsp 279	DB 580 ThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMetLys 599
QY 712 GACACCTCCGATAGTGACAGCGATGCTGACCTTGAAAGAGCTTCAGAACTCGCCCTCGC 771	QY 1597 ATGAAGCCAGAGAGCCCGAGTCTCCGAAATCGCGTGTCTCAG----- 1638
DB 280 AspThrSerAspSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeuArg 299	DB 600 MetLysProGluLysProSerLeuProLysCysGlyAlaGlnProValGlnTrpLeuLeu 619
QY 772 AGAAGCGGAAACAGGAGCCCATGAGAGCCCTCGAAAGGATCCAGAGTCCCTCGCGG 831	QY 1639 -----GCAGGTATTAATAATCTCTCTGTGTCACCAAG 1668
DB 300 ArgLysArgGluGlnGluProThrGluArgProLeuLysGlyIleGlnSerArgLeuArg 319	DB 620 LeuLeuSerSerTrpGlnLeuGlyThrSerSerGlyLeuGlnVal---ThrLeuCysGln 638
QY 832 AAGAGCGCGGAGGAGGTCCCGCGAGCTGCTGCGCTCCGAGGCGCAGTGAACACTGTG 891	QY 1669 AGACCACTCCAGAAAAAAGAGACCCAGTGAAGAGGCGAGTGTGTCTCCTCGCGCG 1728
DB 320 LysLysArgArgGluGluGluProAlaGluThrValGlySerGluAlaSerAspThrVal 339	DB 639 ProProPheThrGluHisLys---TrpLeuLysPheSerValLeu-----His 653
QY 892 GAGGCGCTCTCGCCAGTAAGCAGGAGCCCGAGACGATCAGGGGTTGTGTCACAGCT 951	QY 1729 AGTGAAGCACTCGGGAAGGAAGCAGCT---TGTGAGAGCAGCAGCGCTGTGTGGCG 1782
DB 340 GluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGlyValValSerGlnAla 359	DB 654 GlyGluThrAlaGlyProProAlaHisArgIleArgLysLeuAsnLysTyrTrpGly 673
QY 952 GGGAAAGATCAGAGAGTAACTGAGGGAAGCGGCTCAGGACATCAAGATGAG 1011	QY 1783 AGCGATCAAAATTACAAATGCAATGAAAGCAGAAAAAGACTGTGCTCCCTCGCGCTCACTG 1842
DB 360 GlyLysAspAspArgGluSerLysLeuGluGluLysAlaAlaGlnAspIleLysAspGlu 379	DB 674 AlaCysHisHisCysMetProLeuArgProLeu----- 684
QY 1012 GAGCTCGAGACTTGGCCCGACGAGCCTGAATGTGAGGGTTACGACCCCAAGCCCTG 1071	QY 1843 TTGTATAATGTATGATCATCACCTAGGGGTGGCCCTCTGACCCCTCCCGTTCTTTCTGCG 1902
DB 380 GluProGlyAspLeuGlyArgProLysProGluCysGluGlyTyrAspProAsnAlaLeu 399	DB 685 -----ProCysLeuProGlyTrp 690
	QY 1903 ATAGCCATCCCTGGGCTGTCCAGGACTG---GGAGTTGACAGCTTTGTGTTAAGCTGAT 1959
	DB 691 -----ProArgSerSerProSerLeuValGlyAlaGlySerLeu----- 703
	QY 1960 CACAGACACCGCTGACCATCATCAGCGGAGCAGAGCCCATGTCCAGGATGCTCTCTGCT 2019
	DB 704 ProA-GHisAlaValHis-----LeuAsnGluProCysProLeu 716
	QY 2020 GCCCTGTGTCCA 2031

Db	717 AlaileCysPro 720	:::	US-09-787-016A-1 (1-2610) x AAM79235 (1-647)
RESULT 5			
AAM79235			
ID	AAM79235 standard; protein; 647 AA.		
AC	XX		529 ATGCTCTCTCCCTGAGGATTCCTGAGCCACGTCCTCCCGCCACACAGCCCGAG 588
XX	XX		1 MetProValSerLeuGluAspSerGlyGluProThrSerCysProAlaThrAspAlaGlu 20
XX	AAM79235;		589 ACAGCTCCGAGGCGAGCTGGAAAGCGCTTCTGAGACCAAGAGCGGCCCGCCAGTCTGCT 548
XX	06-NOV-2001 (first entry)		21 ThrAlaSerGluGlySerValGluSerAlaSerGluThrArgSerGlyProGlnSerAla 40
XX	Human protein SEQ ID NO 1897.		649 TCCACAGCTGTGAAGAACGACACCGCTCTTCTGAAAGGTGAAGAGGAGGAGGATGACCC 708
XX	Human; cytokine; cell proliferation; cell differentiation; gene therapy;		41 SerThrAlaValLysGluArgProAlaSerSerGluLysValLysGlyGlyAspAspHis 60
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;		709 GATGACCTCCGATGATGACAGCGATGCTGACCTTCTGAAAGAGCTTCTGAAAGAGCTT 768
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;		61 AspAspThrSerAspSerAspSerAspGlyLeuThrLeuLysGluLeuGlnAsnArgLeu 80
XX	nervous system disorder; arthritis; inflammation.		769 CCCAGAGCGGGAACAGGAGCCCTCTGAGAGCGCCCTGAAAGGGATCCAGATGCCCTG 828
OS	Homo sapiens.		81 ArgArgLysArgGluGlnGluProThrGluArgProLeuLysGlyIleGlnSerArgLeu 100
XX	WO200157190-A2.		829 CGGAGAGCGCGGAGGAGGCTCCCGCGAGACTGTGGGCTCCGAGGCCAGTGACACT 888
XX	09-AUG-2001.		101 ArgLysLysArgArgGluGlnGlyProAlaGluThrValGlySerGluAlaSerAspThr 120
XX	05-FEB-2001; 2001WO-US004098.		889 GTGGAGGGGCTCTCTGCCCAGTAAAGCAGGAGCCCGAGACGATCAGGGGCTTGTCTCCAG 948
XX	03-FEB-2000; 2000US-00496914.		121 ValGluGlyValLeuProSerLysGlnGluProGluAsnAspGlnGlyValValSerGln 140
XX	27-APR-2000; 2000US-00560875.		949 GTTGGAAAGATGACAGAGAGTAAAGTGGAGGGAAGCGCGCTCAGACACATCAAGAT 1008
XX	20-JUN-2000; 2000US-00598075.		141 AlaGlyLysAspAspArgGluSerLysLeuGluGlyLysAlaAlaGlnAspIleLysAsp 160
XX	19-JUL-2000; 2000US-00620325.		1009 GAGGAGCTCGAGACTTGGCGGCGAGCCGAGACCTGAAATGTGAGGGTTACGACCCCAACGCC 1068
XX	01-SEP-2000; 2000US-00654936.		161 GluGluProGlyAspLeuGlyArgProLysProGluCysGluGlyLysArgProAsnAla 180
XX	15-SEP-2000; 2000US-00663561.		1069 CTGTATTGCTTTCGCGGCTCAGCTCAACACAGGTTTATGATTTGCTGTGACCCCTGT 1128
XX	20-OCT-2000; 2000US-00693325.		181 LeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCysCysAspArgCys 200
XX	30-NOV-2000; 2000US-00728422.		1129 GAAGATGTTTCATGGCGATTTGTGGGCTTCTGAGGCTCGAGGGAGGCTTTTGGGA 1188
XX	(HYSE-) HYSEQ INC.		201 GluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGlu 220
XX	Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;		1189 AGGAATGGGGAAGACTATATCTCCCAAACTCCCAATTCGCAAGTGCAGATGAGACT 1248
XX	Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;		221 ArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnValGlnAspGluThr 240
XX	Xue AJ, Yang Y, Wejhrman T, Goodrich R;		1249 CATTCGAAACCGCAGATCAGCAGGAGCTAAATGAGACCTGAGATGCTGATGCGACC 1308
XX	WPI: 2001-476283/51.		241 HisSerGluThrAlaAspGlnGlnGluAlaLysTrpArgProGlyAspAlaAspGlyThr 260
XX	N-PSDB; AAKS2368.		1309 GATTGTACAAATATAGCAATAGACAGCAAGTCTAGCGAAGACCAAGCGATTAAGGGT 1368
XX	Nucleic acids encoding polypeptides with cytokine-like activities, useful		261 AspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGly 280
XX	in diagnosis and gene therapy.		1369 AGAATTGAGAAAGCTCCAATCCAGTGCAGAGAAAGAACTCAGATCTTCCAGCT--- 1425
XX	Claim 20; Page 4293-4294; 6221pp; English.		281 ArgIleGluLysAlaAlaAsnProSerGlyLysLysLeuLysIlePheGlnProGly 300
XX	The invention relates to polynucleotides (AAKS1456-AAKS3435) and the		1425 ----- 1425
XX	encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to		301 ProGlyProValProThrGlnLeuProValLeuTrpGlnValLeuGluIleAlaValSer 320
XX	cytokine, cell proliferation or cell differentiation or which may induce		1426 -----GTGATAGAGGCGCCT 1440
XX	production of other cytokines in other cell populations. The		321 ArgSerIleSerAlaPheThrLeuLeuHisCysIleSerCysLysValIleGluAlaPro 340
XX	polynucleotides and polypeptides are useful in gene therapy, vaccines or		1441 GGTGCTCTCAAAATGATTCGGGCTGTGTGAGTGGCGGAGCCCGACCTCGGTGTAC 1500
XX	peptide therapy. The polypeptides have various cytokine-like activities,		341 GlyAlaSerLysCysIleGlyProGlyCysCysHisValAlaGlnProAspSerValTyr 360
XX	e.g. stem cell growth factor activity, haematopoiesis regulating		
XX	activity, tissue growth factor activity, immunomodulatory activity and		
XX	activin/inhibin activity and may be useful in the diagnosis and/or		
XX	treatment of cancer, leukaemia, nervous system disorders, arthritis and		
XX	inflammation. Note: Records for SEQ ID NO 2110 (AAKS2581), 2111		
XX	CC (AAKS2582) and 3666 (AAM80020) are omitted as the relevant pages from the		
XX	sequence listing were missing at the time of publication		
XX	Sequence 647 AA;		

Alignment Scores:			
Pred. No.:	2,37e-171	Length:	647
Score:	2311.00	Matches:	441
Percent Similarity:	92.45%	Conservative:	0
Best Local Similarity:	92.45%	Mismatches:	36
Query Match:	49.29%	Indels:	36
DB:	4	Gaps:	1

QY 1501 TGCAGTAATGACTGATCTCTCAACACGCGCAGCGACAAATGAAGTTTCTTAAGCTCAGGT 1560
 Db 361 CysSerAspCysIleLeuYsHsAlaAlaAlaThrMetLysPheLeuSerGly 380
 QY 1561 AAAGAACAAGACCCAAAGCCTAAAGAAAGATGAAGATGAAGCCAGAGCCAGCTCTT 1620
 Db 381 LysGluGlnLysProLysProLysGluLysMetLysMetLysProGluLysProSerLeu 400
 QY 1621 CCGAATGCGGTCTCAGCAGCTATTAATACTCTTCTGTGCACAAAGAGACCACTCCA 1680
 Db 401 ProLysCysGlyAlaGlnAlaGlyIleLysIleSerSerValHisLysArgProAlaPro 420
 QY 1681 GAAAAAAGAGACCACTGAGCAAGAGCAGTGTGTCTCTCCGCGGAGTCAAGCACTC 1740
 Db 421 GluLysLysGluThrThrValLysLysAlaValValProAlaArgSerGluAlaLeu 440
 QY 1741 GGAAGCAAGACCTGTGTGAGCAGCAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800
 Db 441 GlyLysGluAlaAlaCysGluSerSerThrProSerThrProSerThrProSerThrPro 460
 QY 1801 GCAGTAAAGCCAGAAAGACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1851
 Db 461 AlaValLysProGluLysThrAlaAlaProSerProSerLeuLysLys 477
 RESULT 6
 ID AAY67580
 AC AAY67580 standard; protein; 614 AA.
 XX AAY67580;
 DT 19-JUN-2000 (first entry)
 DE Murine death inducer-obliterator 1 (DIO-1) polypeptide.
 KW Death inducer-obliterator 1; DIO-1; cell death; cancer; tumour; mouse;
 KW autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
 KW antirheumatic; antiinflammatory; antiproliferative.
 OS Mus sp.
 XX Key
 FH Region
 FT Location/Qualifiers
 FT 162..170
 FT /note= "NLS sequence"
 FT 182..190
 FT /note= "NLS sequence"
 FT 262..285
 FT /note= "zinc finger motif"
 FT 290..315
 FT /note= "zinc finger motif"
 XX WO200015787-A1.
 XX 23-MAR-2000.
 XX 10-SEP-1999; 99WO-GB003019.
 XX 10-SEP-1998; 98SE-00003069.
 XX 17-SEP-1998; 98US-0100873P.
 XX (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
 XX (BANN/) BANNERMAN D G.
 XX Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
 XX WPI, 2000-271426/23.
 XX N-PSDB; A4290579.

New DNA encoding human and murine death inducer-obliterator 1 polypeptides, useful in the treatment of cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumors, malignant tumors and hyperproliferative skin disorders.

Claim 8; Fig 1D; 27pp; English.

The invention provides nucleic acids encoding the human and murine death inducer-obliterator 1 (DIO-1) polypeptides. The polypeptides can be expressed by standard recombinant methodology. The DIO-1 polypeptides, agonists and antagonists are used as a medicament for treating diseases characterized by an alteration in cell death or by hyperproliferation, e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign tumours, malignant tumours or hyperproliferative skin disorders. They are also useful in the treatment of metabolic, proliferative or inflammatory conditions. The present sequence represents the murine DIO-1 polypeptide

Sequence 614 AA;

Alignment Scores:
 Pred. No.: 3,79e-159 Length: 614
 Score: 2154.50 Matches: 449
 Percent Similarity: 76.96% Conservative: 42
 Best Local Similarity: 70.38% Mismatches: 112
 Query Match: 45.95% Indels: 36
 Gaps: 12

US-09-787-016A-1 (1-2610) x AAY67580 (1-614)

QY 265 AFGAGACGACAAAGCGACCCCGAGCAATGAGGAGCACCCTAAGGCCATCAAAACCCACGAC 324
 Db 1 MetAspAspLysGlyHisLeuSerAsnGluGluAlaProLysAlaIleLysProThrSer 20
 QY 325 AAAGAGTTCCAGAAACATGGGTTTCGAAAGGACCACTATCGCCAGCGAGAGCGCGCA 384
 Db 21 LysGluPheArgLysThrTropGlyPheArgArgThrIleAlaLysArgGluGlyAla 40
 QY 385 GGGAGCGGCGGAGCTGACCCACTGAGCGCCGCCACCCACACGACGACGAGCTGGCGCTGTCC 444
 Db 41 GlyAspThrGluAlaAspProSerGluGlnGlnPro-----GlnGlnHisAsnLeuSer 58
 QY 445 CTGCGCGCAGTGGGAGGAGCGCCAAAGCCACTGAGCGGTGGAGCAGCTTCCTGACCAT 504
 Db 59 LeuArgArgSerGlyArgGlnProLysArgThrGluArgValGluGluPheLeuThrThr 78
 QY 505 GCGCGCGCGCGGAGGAGGAGCATGCTGCTCTCTGAGGATTTCTGTGAGCCACG 564
 Db 79 ValArgArgArgGlyLysLysAsnValProValSerLeuGluAspSerSerGluProThr 98
 QY 565 TCTGTCCCGCCGACAGACCGCAGACAGCTCCGAGGCGGAGGTGAGAAAGCGCTTCTGAG 624
 Db 99 SerSerThrValThrAspValGluThrAlaSerGluGlySerValGluSerSerSerGlu 118
 QY 625 ACCAGAGCGGCGCCCGACGCTCTCTCCACAGCTGTGAGGAAACGACGACGCTCTTCTGAA 684
 Db 119 IleArgSerGlyProValSerAspSerLeuGly---LysGluHisProAlaSerSerGlu 137
 QY 685 AAGGTGAAGAGGAGGAGTACCAACGATGACACTCTCGATGATGACAGCATGCGCTGACC 744
 Db 138 LysAlaLysGlyGlyGluGluGluAspThrSerAspSerAspSerAspGlyLeuThr 157
 QY 745 TTGAAGAGCTTCAGATCGCTTCGACAGAGCGGAAACAGGAGCCACTGAGAGCGCC 804
 Db 158 LeuLysGluLeuGlnAsnArgLeuArgLysLysArgGluGluAspSerAlaGluThr 177
 QY 805 CTGAAGAGGATCCAGAGTGCCTCTGCGGAAGAACCGCGGAGGAGGGTCCCGCGGAGCT 864
 Db 178 LeuArgGlySerGlnAsnArgLeuArgLysLysArgGluGluAspSerAlaGluThr 197
 QY 865 GTGGGCTCCGAGCCAGTGCATCTGTGGGGGGCTCTCTGCC---AGTAGCAGGAGCGCC 921
 Db 198 ---GlySerValGlnIleGlySerAlaGluGlnAspArgProLeuCysLysGlnGluPro 216
 QY 922 GAGAACGATCAGGGGGTGTGTCTCCAGGCTGGGAAAGATGACAGAGACTAAGTTGGAG 981
 Db 217 GluAlaSerGlnGlyProValSerGlnSerGluThrAspAspIleGluAsnGlnLeuGlu 236
 QY 982 GGAAGCGCGCTCAGGACATCAAGATGAGAGCCTGGAGACTTGGCCCGACCGACGCT 1041

Query Match:	33.08%	Indels:	36
DB:	4	Gaps:	1
US-09-787-016A-1 (1-2610) x ABG22388 (1-330)			
Qy	967	GAGAGTAAGTTGAGGGAAGGCGGCTCAGGACATCAAGATGAGGAGCTGAGACTTG	1026
Db	1	GluserLysLeuGluGlyLysAlaAlaGlnAspIleLysAspGluGluProGlyAspLeu	20
Qy	1027	GGCGACCGAGCTGAGTGGTGTACGACCCCAAGCCCTGATTGTCATTTCGCCG	1086
Db	21	GlyArgProLysProGluCysGluGlyTyrAspProAsnAlaLeuTyrCysIleCysArg	40
Qy	1087	CAGCCTCACACAAACAGTTTATGATTGCTGACCGCTGTGAAGATGGTTTCAGGC	1146
Db	41	GlnProHisAsnAsnArgPheMetIleCysCysAspArgCysGluGluTyrPheHisGly	60
Qy	1147	GATTGTGGGCATTCTGAGGCTCGAGGAGCTTTTGCAAGGATGGGAGACTAT	1206
Db	61	AspCysValGlyIleSerGluAlaArgGlyArgLeuLeuGluArgAsnGlyGluAspTyr	80
Qy	1207	ATCTGCCCAACTGCACCATCTGCAAGTCAGGATGAGACTCATTTCAGAAACGGCAGAT	1266
Db	81	IleCysProAsnCysThrIleLeuGlnValGlnAspGluThrHisSerGluThrAlaAsp	100
Qy	1267	CAGCAGAACTAAATGAGACTGGAGATGCTGATGCGACCGATGTACAGTATAGGA	1326
Db	101	GlnGlnGluAlaLysTyrArgProGlyAspAlaAspGlyThrAspCysThrSerIleGly	120
Qy	1327	ACAATGAGCAGAGTCTAGCAGACCAAGGATTAAGGCTAGAAATTGAGAAAGCTGCA	1386
Db	121	ThrIleGluGlnLysSerSerGluAspGlnGlyIleLysGlyArgIleGluLysAlaAla	140
Qy	1387	AATCAAGTGCAGAGAAACTCAAGATCTTCAGCCT	1425
Db	141	AsnProSerGlyLysLysLysLeuLysIlePheGlnProGlyProGlyProValProThr	160
Qy	1425	-----	1425
Db	161	GlnLeuProValLeuTyrGlnValLeuGluIleAlaValSerArgSerIleSerAlaPhe	180
Qy	1426	-----GTGATGAGCGCGCTGGTGCCTCAAAATGATT	1458
Db	181	ThrLeuLeuHisCysIleSerCysLysValIleGluAlaProGlyAlaSerLysCysIle	200
Qy	1459	GGCCCCGGTCTGTCTGCTGGCGCAGCCGACTCGGTACTGCTGCTGCTGCTGCTGCT	1518
Db	201	GlyProGlyCysCysHisValAlaGlnProAspSerValTyrCysSerAsnAspCysIle	220
Qy	1519	CTCAACACCGCGCAGCAGCAATGAAGTTCTAAGCTCAGGTAAAGAACAGACCAAG	1578
Db	221	LeuLysHisAlaAlaAlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLys	240
Qy	1579	CCTAAGAAAGATGAGATGACCCAGAGAGCCCGCTCTCCGAATCGGTGCTCAG	1638
Db	241	ProLysGluLysMetLysMetLysProGluLysProSerLeuProLysCysGlyAlaGln	260
Qy	1639	GCAGTATTAATCTCTTCTGTCACAGAGACAGCTCCAGAAAAAAGACACACA	1698
Db	261	AlaGlyIleLysIleSerSerValHisLysArgProAlaProGluLysLysGluThrThr	280
Qy	1699	GTGAAGAGGAGGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1758
Db	281	VallLysLysAlaValValProAlaArgSerGluAlaLeuGlyLysGluAlaAlaCys	300
Qy	1759	GAGAGCAGCGCTGTGTGGCGGCGGATCACATTAATGAGTAAAGCCAGAAAG	1818
Db	301	GluserSerThrProSerTrpAlaSerAspHisAsnTyrAsnAlaValLysProGluLys	320
Qy	1819	ACTGCTGCTCCCTCGCCG	1836
Db	321	ThrAlaAlaProSerPro	326

RESULT 8	
AAB43724	
ID	AAB43724 standard, protein; 181 AA.
XX	
AC	AAB43724;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human cancer associated protein sequence SEQ ID NO:1169.
XX	
KW	Human; cancer associated gene; cancer antigen; detection; cancer;
KW	diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW	antiidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW	antiinflammatory; antithyroid; antiallergic; antibacterial; cardiac;
KW	dermatological; neuroprotective; thrombolytic; coagulant; neoplastic;
KW	vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW	immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW	allergic reaction; graft versus host disease; organ rejection;
KW	haemostatic; thrombolytic; cardiovascular disorder; infection;
KW	neurological disease; drug screening.
XX	
OS	Homo sapiens.
XX	
PN	WO200005350-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US005882.
XX	
PR	12-MAR-1999; 99US-0124270P.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI; 2000-587533/55.
DR	N-PSDB; AAC77933.
XX	
PT	Novel isolated nucleic acids comprising sequences encoding peptides
PT	useful for treating or diagnosing e.g. cancer.
XX	
PS	Claim 11; Page 1790-1791; 2352pp; English.
XX	
CC	AAC77607 to AAC78448 encode the human cancer associated proteins given in
CC	AAB43398 to AAB44239. The proteins can have activities based on the
CC	tissues and cells the genes are expressed in. Example of activities
CC	include: cytostatic; proliferative; vulnerary; immunomodulator;
CC	antiidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC	antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC	dermatological; neuroprotective; cardiac; thrombolytic; coagulant;
CC	neoplastic; vasotropic; antipsoriatic and antiangiogenic. The
CC	polynucleotides and polypeptides can be used for preventing, treating or
CC	ameliorating medical conditions and diagnosing pathological conditions.
CC	Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC	the present invention may be used to treat immune disorders by activating
CC	or inhibiting the proliferation, differentiation or mobilisation of
CC	immune cells, to treat disorders of haematopoietic cells, autoimmune
CC	disorders, allergic reactions, graft versus host disease, and organ
CC	rejection, modulate haemostatic or thrombolytic activity, modulate
CC	inflammation, cancers, cardiovascular disorders, neurological disease and
CC	bacterial or viral infections. The peptides, nucleotides, antibodies,
CC	agonists and antagonists may be also be used in drug screens. AAC78449 to
CC	AAC78457 and AAB44240 represent sequences used in the exemplification of
CC	the present invention
XX	
SQ	Sequence 181 AA;
Alignment Scores:	
Pred. No.:	5,63e-66 Length: 181
Score:	98.00 Matches: 179
Percent Similarity:	98.35% Conservatives: 0
Best Local Similarity:	98.35% Mismatches: 2
Query Match:	20.43% Indels: 2

```

DB:          3          0          Gaps:
US-09-787-016A-1 (1-2610) x ABA43724 (1-181)
QY 998 ACATCAAGATGAGGAGCTGGAGACTTG-GCCGACCGAACCCCTGAATGTGAGGTTAC 1056
Db 1 ThrSerLysMetArgSerLeuGluThrLeuGlyArgProLysProGluCysGluGlyTyr 20
QY 1057 GACCCCAAGCCCTGATGCTATGTCGCGCCAGCTCACAAACAGGTTTATGATTGTC 1116
Db 21 AspProAsnAlaLeuTyrCysIleCysArgGlnProHisAsnAsnArgPheMetIleCys 40
QY 1117 TGTGACCGCTGCAAGAAATGTTTCATGCGCGATTGTGTGGCAATTCGTAGGCTCGAGG 1176
Db 41 CysAspArgCysGluGluTrpPheHisGlyAspCysValGlyIleSerGluAlaArgGly 60
QY 1177 AGCTTTTGGAAAGGATGGGAGACTATATCTGCCAACTGCACCACTTCCTGCAAGTG 1236
Db 61 ArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGlnVal 80
QY 1237 CAGGATGAGACTCATTCAAGAACGGCAGATCAGCAGGAGCTAAATGGAGACTGGAGAT 1296
Db 81 GlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaIleTyrArgProGlyAsp 100
QY 1297 GTGTATGGACCCGATGTACAGTATAGGAACATAGACAGAGTCTAGCGAGACCAA 1356
Db 101 AlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAspGln 120
QY 1357 GGGTAAAGGGTAGATTGAGAAAGCTGCAAAATCCAAATGCGCAAGAGAACTCAAGATC 1416
Db 121 GlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLysIle 140
QY 1417 TTCCAGCTGTGATAGAGCGCTGTGTGCTCAAAATGTATTGGCCCGGCTGTCTAC 1476
Db 141 PheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysHis 160
QY 1477 GTGGCCGACCCGACTCGGTGACTGTCAGTATGACTGTATCTCAACACCGCCGAGG 1536
Db 161 ValAlaHis-ProThrArgCysThrAlaValMetThrValSerSerAsnThrProGlnAr 180
QY 1537 ACAA 1540
Db 180 gGln 181
RESULT 9
ID ABG22387
XX ABG22387 standard; protein; 167 AA.
XX AC
XX ABG22387;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #22378.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US008631.
XX PR 31-MAR-2000; 2000US-00540217.
XX PR 23-AUG-2000; 2000US-00649167.
XX PA (HYSB-) HYSBQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR

```

```

DR N-PSDB; AAS86574.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
PS Claim 20; SEQ ID NO 52746; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
sequences. (I) is useful as hybridisation probes, polymerase chain
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
and in recombinant production of (II). The polynucleotides are also used
in diagnostics as expressed sequence tags for identifying expressed
genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 167 AA;
SQ
Alignment Scores:
Pred. No.: 2,636-61 Length: 167
Score: 898.00 Matches: 167
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 19.15% Indels: 0
DB: Gaps: 4
US-09-787-016A-1 (1-2610) x ABG22387 (1-167)
QY 1174 GGGAGGCTTTTGGAAAGGAATGGGGAAGACTATATCTGCCAAAGCTGACCACCTTCGCAA 1233
Db 1 GlyArgLeuLeuGluArgAsnGlyGluAspTyrIleCysProAsnCysThrIleLeuGln 20
QY 1234 GTCCAGGATGAGACTCATTTCAGAAACGGCAGATCAGCAGGAGCTTAATGGAGACTGGA 1293
Db 21 ValGlnAspGluThrHisSerGluThrAlaAspGlnGlnAlaLysTrpArgProGly 40
QY 1294 GATGCTGATGGCAGCGATTGTACAGTATAGGAACAATAGACAGCAAGTCTAGCGAAGAC 1353
Db 41 AspAlaAspGlyThrAspCysThrSerIleGlyThrIleGluGlnLysSerSerGluAsp 60
QY 1354 CAAGGGAATAAGGGTAGAATTGAGAAAGCTGCAAAATCCAAATGCGCAAGAGAACTCAAG 1413
Db 61 GlnGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLysLysLysLys 80
QY 1414 ATCTTCCAGCTGTGATAGAGCGCTGTGTGCTCAAAATGTATTGGCCCGGCTGTCT 1473
Db 81 IlePheGlnProValIleGluAlaProGlyAlaSerLysCysIleGlyProGlyCysCys 100
QY 1474 CACGTGGCGCAGCCGACTCGGTGTACTGTCAGTATGACTGTATCTCAAAACGCGCA 1533
Db 101 HisValAlaGlnProAspSerValTyrCysSerAsnAspCysIleLeuLysHisAla 120
QY 1534 GCGACAATGAAGTTTCTAAGCTCAGGTAAAGAACAGAGCCAAAGCTTAAGAAAGATG 1593
Db 121 AlaThrMetLysPheLeuSerSerGlyLysGluGlnLysProLysProLysGluLysMet 140
QY 1594 AAGATCAAGCCAGAGAGCCAGTCTTCGGAATCGGTGCTCAGCAGGTATTAAATC 1653

```

Db 141 LysMetLysProGluLysProSerLeuProLysCysGlyAlaGlnAlaGlyIleLysIle 160
QY 1654 TCTTCTGTGCAAGAGACCA 1674
Db 161 SerSerValHisLysArgPro 167
RESULT 10
ABP05496
ID ABP05496 standard; protein; 108 AA.
XX
AC ABP05496;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:10974.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI: 2002-106308/14.
DR N-PSDB; ABN21248.
XX
XX Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
XX
PS Disclosure; SEQ ID NO 10974; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage. N.B. The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 108 AA;
Alignment Scores:
Pred. No.: 7,52e-30 Length: 108
Score: 494.00 Matches: 94
Percent Similarity: 89.81% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 11
Query Match: 10.54% Indels: 0
DB: 5 Gaps: 0
US-09-787-016A-1 (1-2610) x ABP05496 (1-108)
QY 1342 TCTAGCGAGACCAAGGATTAAGGTTGAGAAAGCTGCAATCCAAATCCAGTGGCAAG 1401
Db 1 SerGlyGluAspHisGlyIleLysGlyArgIleGluLysAlaAlaAsnProSerGlyLys 20
QY 1402 AAGAACTCAAGATTTCCAGCTGTGTAGAGGCGCTGTGCTCAAAATGATTTGGC 1461
Db 21 LysLysLeuLysIlePheGlnProValGluAlaProGlyAlaProLysCysIleGly 40
QY 1462 CCGGGTGTCTGTCAGTGGCGCAGCCGACTCGGTGTACTGCAGTAATGACTGTATCCTC 1521
Db 41 ProGlyCysSerSerValAlaGlnProAspSerValLysCysSerAsnAspCysIleLeu 60
QY 1522 AAACAGCGCGCAGCACAATCAAGTTTCTAAGCTCAGTAAAGAACAGACCAAGCCT 1581
Db 61 LysHisAlaAlaAlaThrMetArgPheLeuSerSerGlyLysGluGlnLysThrLysPro 80
QY 1582 AAAGAAAGATGAGATGAGCCAGAGAGCCAGTCTCCGAAATGCGGTGCTCAGGCA 1641
Db 81 LysGluLysValLysThrLysProGluLysPheSerLeuProLysCysSerValGlnVal 100
QY 1642 GGTATTAATCTCTCTGTGCAC 1665
Db 101 GlyIleLysIleSerSerValHis 108
RESULT 11
ABP71487
ID ABP71487 standard; protein; 2016 AA.
XX
AC ABP71487;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 41253.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKS) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
DR N-PSDB; ABL15590.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
PS Disclosure; SEQ ID NO 41253; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 2016 AA;

Alignment Scores:
 Pred. No.: 9,27e-17 Length: 2016
 Score: 332.00 Matches: 159
 Percent Similarity: 32.52% Conservative: 68
 Best Local Similarity: 22.78% Mismatches: 221
 Query Match: 7.08% Indels: 250
 DB: 4 Gaps: 27

US-09-787-016A-1 (1-2610) x ABB71487 (1-2016)

QY 256 GGAGCAGGTATGACGACAAAGGCGACCGCAGCAATGAGGAGCAGCTAAGGCCATCAAA 315
 DB 676 GlyValGlyAsnAlaAlaAlaAspProAsn-----IleMet 689
 QY 316 CCCACGACAAAGAGTTTCAGAAACATCGGGTTTCGAAAGACCACTATCGCCAGCGA 375
 DB 690 AspThrAlaAsnGluAspGluIleThrAlaAspPheLeuGlnHisValValGlyLeuIle 709
 QY 376 GAGGGCGGAGCGCGGAGGTGACCCACTGGAGCGCGCCACCCACAGCAGCAGCTG 435
 DB 710 GluGluAspLysGlnPheGluAla----- 717
 QY 436 GGCCTGTCTCCGCGCGCAGTGGGAGCAGCCCAAGCGCACTGAGCGCTGGAGCAGTTC 495
 DB 718 -----GluValValLysGlnVal 723
 QY 496 CTGACACMTCCGCGCGCGCGGAGGAGGAGCATGCTGTCTCCCTGGAGGATCTGGT 555
 DB 724 LeuAla---SerThrGluProGlyThrLeuAspAlaIleValSerMetProThrSerIle 742
 QY 556 GAGCCACGCTCTCCCGCCGACAGAGCGCGGAGCAGCTCCGAGGCGACGTG---GAA 612
 DB 743 GluProValAspValProGlnAlaHisThrAsnLeuLeuProAsnAlaSerLeuThrGlu 762
 QY 613 AGCGCTTCTGAGACGAGAGCGCGCCAGTCTGCTTCCACA-----GCT 657
 DB 763 ProAlaGlnSerMetThrSerLeuProIleAlaCysSerThrProSerArgSerValAla 782
 QY 658 GTGAGGACGACGACGCTCTCTGAAAGGTGAAGAGGGGATGACACGATGACACC 717
 DB 783 AlaSerThrProProThrSerAlaLysValValArgly----- 795
 QY 718 TCCGATAGTGACAGCGATGCGCTGACCTTCAAAGAGCTTCAGAAATCGCTCCGAGGAAG 777
 DB 796 -----TyrGlyArgVal 799
 QY 778 CGGGAACAGAGAGCCACTGAGAGCCCTGAAAGGATCCAGATCCCTCGCGAAGAG 837
 DB 800 IleTyrLeuProIleGluAlaPro-----ThrThrArgAlaLys 813
 QY 838 CGCCGGGAGGAG-----GGTCCCGCGAGACTGTGGGCTCCGAGGCC--- 879
 DB 814 ArgArgAlaGlnPheProSerAlaProGlyMetAlaAlaThrSerSerSerAspAlaGly 833
 QY 879 ----- 879
 DB 834 AsnLeuSerPheGlyGluSerSerLeuAspAlaSerIleAsnGlnProLeuAsnThrSer 853
 QY 880 -----AGTGACACTGTGGGGCGCTCTGCGCCAGTAAGCAGGAGCGCGGAGACGAT 930

DB 854 SerLeuSerAsnAspSerGlnProGlySerGlyProLysArgProAsnPro---ArgGlu 872
 QY 931 CAGGGGGTGTGTCCTCCAGCTCGGAAAGATGACAGAGAGTAAAGTTCGAGGAAAGCG 990
 DB 873 ProSerMetAlaArgArgSerThrAlaProArgArgSerLysLysLeuAspAlaSerGln 892
 QY 991 GTCAGGAC-----ATCAAAGATGAGGAGCTGGAGACTTGGCCGACCGAGCCT 1041
 DB 893 AsnAsnAspProAspAlaSerGlnSerGlnGluAspAsp----- 906
 QY 1042 GAATGTGAGGGTTACGACCCCAACGCCCTGTATTGCATTTCGCCGAGCCTCACACAAC 1101
 DB 907 -----AspProAsnLysLeuTyrCysIleCysArgGlnProHisAsnAsn 921
 QY 1102 AGTTTATGATTGCTGTGACCCGCTGTGAGAAATGGTTTCATGCGGATTTGTGGGCATT 1161
 DB 922 ArgPheMetIleCysCysAspLeuCysGluAspTyrPheHisGlyThrCysValGlyVal 941
 QY 1162 TCTGAGGCTCGAGGAGGCTTTTGAAGAGGAATGGGAGAGCTATATCTGCCCAACTGC 1221
 DB 942 ThrLysAlaMetGlyThrAspMetGluAsnLysGlyIleAspTyrLysCysProLysCys 961
 QY 1222 ACCATTCTCAAGTCGAGATGAGACTCATTCAAGAAACGCGCAGATCAGCAGAGATAA 1281
 DB 962 -----ValLysArgGlnGluArg 968
 QY 1282 TGGAGACCTGGAGATGCTGATGCGACCGATTGTACAAATATAGCAATATAGACGAGAAG 1341
 DB 969 SerGlnProArgIleThrAspMetLeuValThrArgProThrThrGlnProGluGlnArg 988
 QY 1342 TCTAGCAGACCAAGGGATA-----AAGGTGAGATTGAGAAAGCTCGAATCAAGT 1395
 DB 989 ProSerGluThrLysValLeuThrThrAlaGluIleValGlnValAlaAlaProSer 1008
 QY 1396 GGC---AAGAAGAACTCAAGATCTCCAGCTGTGTATAGAGCGCT----- 1440
 DB 1009 AlaProArgArgThrLeuProValValLeuThrValAlaSerSerProMetArgIlePro 1028
 QY 1441 -----GGTGCCTCAAA----- 1452
 DB 1029 MetAlaLysProAlaLysLysPheProThrGlyAlaIleSerHisGlnGlnGlnGln 1048
 QY 1453 -----TGATTGCGCCCGGG----- 1467
 DB 1049 LeuAsnPheIleArgLeuGlyProSerProGlyLysArgIleSerGluThrLeuCysVal 1068
 QY 1468 ---TGCTGTACGTGCGCGCAGCGCTCGGTACTGCTGAGTAATGACTGTATCTCTCAA 1524
 DB 1069 ValCysLysArgProAlaSerThrSerSerValTyrCysGlyGluCysIleArgLys 1088
 QY 1525 CACGCC-----GCA 1533
 DB 1089 TyrAlaGlnSerAlaIleGlnAlaHisAlaAlaThrLysGlyProLeuProGlnAsnAla 1108
 QY 1534 GGCACATCAAGTTCTAAGCTCAGCTAAAGACAGACAGCCAAAGCCTAAAGAAAAGATG 1593
 DB 1109 GlyAlaGlnSerLeuLeuAsnAsnSerPheAspAlaLysAsnLysLysLysAspLeu 1128
 QY 1593 ----- 1593
 DB 1129 PheGluAspValLeuArgGlnAlaAspThrValSerLysValGluArgIleAsnValPhe 1148
 QY 1593 ----- 1593
 DB 1149 GluArgLysSerGlyArgValIleThrGlyHisMetAlaProSerSerAlaHisGlnPheArg 1168
 QY 1594 AAGATCAAGCCAGAGAGAGCCAGT-----CTTCCGAAATCGCGTCTCAGGAGGCT 1644
 DB 1169 LysTyrLeuGlnGluAsnProSerPheGluValLeuProSerGlyThrValGlnSer--- 1187
 QY 1645 ATTAATAATCTTCTTGTGCGACAGAGA-----CCAGCTCCAGAAAAAAGAGACC 1695
 DB -----

Db 1188 -----AlaAspAlaGluLysArgLeuLeuLysGlyAlaProGluAlaAlaThrSer 1204
 Qy 1696 ACAGTGAAGAGGACGTGGTGGTCCCTGGCGGAGT-----GAAGCACTGGGAGGAA 1749
 Db 1205 ThrSerGluProAlaValLeuGlyValAlaLysLysProGluGlyProAlaLysLeu 1224
 Qy 1750 GCAGCTTGTGAGACGACGCGCTGGTGGCGGCGGATCACAATTACAAATGAGTAAG 1809
 Db 1225 SerHisProGlnAsnThrThrValGlnAlaSerHisGlnLeuGlyIleSerSerValArg 1244
 Qy 1810 CCA-----GAAAGACGTCT-----CCTCCCTCGCGGTCA 1839
 Db 1245 ProLeuAlaLysLysAspLysGluLysThrProThrValGlnAlaProThrProAsn 1264
 Qy 1840 CTGTTGTATAATGTATGATACCTAGGCGTTGGCTCTCGGACCCCTCCCGT 1893
 Db 1265 -----ArgIleAlaAlaGlyLysProGluProValArg 1275
 RESULT 12
 AAM78908
 ID AAM78908 standard; protein; 1539 AA.
 XX
 AC AAM78908;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 1570.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WC200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001W0-US004098.
 XX
 PR 03-FEB-2000; 2000US-00496914.
 PR 27-APR-2000; 2000US-00560875.
 PR 20-JUN-2000; 2000US-00598075.
 PR 19-JUL-2000; 2000US-00620325.
 PR 01-SEP-2000; 2000US-00654936.
 PR 15-SEP-2000; 2000US-00663561.
 PR 20-OCT-2000; 2000US-00693325.
 PR 30-NOV-2000; 2000US-00728422.
 XX
 PA (HVSE-) HVSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
 PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren P, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 WPI; 2001-476283/51.
 DR N-PSD3; AAK52041.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
 PT in diagnosis and gene therapy.
 XX
 PS Claim 20; Page 3896-3899; 6221pp; English.
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity relating to
 CC cytokine cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
 CC sequence listing were missing at the time of publication
 XX
 SQ Sequence 1539 AA;
 Alignment Scores:
 Pred. No.: 5,35e-09 Length: 1539
 Score: 232.00 Matches: 94
 Percent Similarity: 36.43% Conservative: 63
 Best Local Similarity: 21.81% Mismatches: 154
 Query Match: 4.95% Indels: 120
 DB: 4 Gaps: 17
 US-09-787-016A-1 (1-2610) x AAM78908 (1-1539)
 Qy 538 TCCTTGGAGGAT---TCGTGTGAGCCACGCTCTCCCGCCACAGACGCGGAGACGCC 594
 Db 95 SerLeuSerAspLysSerHisAlaHisProGlyCysLeuLysGluProHisProAla 114
 Qy 595 TCCGAGGCGAGGTGGAAGGCGCTTCTGAGACCAGA---AGCGGCCCGCAGTCTGCTCC 651
 Db 115 GlnThrGlyHisValSerHisSerGlnLysGlnCysHisLysProGlnGlnAla 134
 Qy 652 ACAGCTGTGAAGAACGACGACGCTCTTCTGAAAAGTGAAA---GGAGGGGATGACCCAC 708
 Db 135 ProAlaMetLysThrAsnSerHisValLysGluGluLeuHisProGlyValGluHis 154
 Qy 709 GATGACATCTCGATAGTACAGCATGCGCTGACCTTGAAGAG---CTT 756
 Db 155 -----PheLysGluLysAspLysLeuLysLysLysLysLysLysLysLysLysLys 170
 Qy 757 CAGAAATCGCTTCGCGAGGAGCGGAAACAGGAGGCCACTGAGAGCGCCCTGAAAGGGATC 816
 Db 171 GlnProArgGlnArgArgSerSerLysSerPheSerLeuAspGluProProLeuPheIle 190
 Qy 817 CAGAGTCGCTCGGAGAGAGCGCGCGGAGAGGCTCCCGCCGAGACTGTGGCTCCGAG 876
 Db 191 ProAspAsnIleAlaThrIleArgArgGlu-----GlySer--- 202
 Qy 877 GCCAGTGACACTGTGGAGGGGCTCTGCCCTAGCAGGAGCGCGGAGAACGATCAGGGG 936
 Db 202 ----- 202
 Qy 937 GTTGTGTCCAGGCTGGGAAAGATCAGAGAGAGTAACTTGGAGGAAAGCGGCTCAG 996
 Db 203 -----AspHisSerSerSerPheGluSerLysTyr----- 212
 Qy 997 GACATCAAGATGAGGAGCGCTGAGACTTGGCGCCGAGCGCCCTGAATGTGAGGGTTAC 1056
 Db 213 -----MetThrProSerLysGlnCys---GlyPhe 222
 Qy 1057 GACCCCAAGCCCTGTATTGCAATTTGCGCCGAGCCCTCACAACAGAGTTTATGATTGC 1116
 Db 223 -----CysLysLysProHisGlyAsnArgPheMetValGly 234
 Qy 1117 TGTGACCGCTGTGAAGAATGGTTTATCGCGATTTGTGGCATTTCTGAGCTCCGAGG 1176
 Db 235 CysGlyArgCysAspAspTrpPheHisGlyAspCysValGlyLeuSerLeuSerGlnAla 254
 Qy 1177 AGCGCTTTTGGAAAAGAAATGGGGAAGACTATCTGCCAAACTGACCACTTCTGCAAGTG 1236
 Db 255 GlnGlnMetGlyGluGluAspLysGluTyrValCysValLysCysCys-----Ala 271
 Qy 1237 CAGGATGAGACTCATTTGAGAAACGGCAGAT-----CAGCAGGAGGCT 1278
 Db 272 GluGluAspLysLysThrGluIleLeuAspProAspThrLeuGluAsnGlnAlaThrVal 291
 Qy 1279 AAATGAGACCTGGAGATCTGATGCGACCGGATTGTACAAGTATAGGAACAATAGAGCAG 1338
 Db 292 GluPheHisSerGlyAspLysThr---MetGluCysGluLysLeuGlyLeuSerLysHis 310


```

Db 412 -----CysLysLysProHisGlyAsnArgPheMetValGly 423
|||.....|
1117 TGTGACGCTGTGAAGATGTTTCATGGCGATGTGTGGCATTTCTGAGGCTCAGGG 1176
|||.....|
Db 424 CysGlyArgCysAspSerTrpPheHisGlyAspCysValGlyLeuSerLeuGlnAla 443
|||.....|
1177 AGCTTTTGGAAAGTAATGGGAGACTATATCTGCGCAAACTGCACCACTTCTGCAAGTG 1236
|||.....|
Db 444 GlnGlnMetGlyGluAspLysGluTyValCysValLysCysCys-----Ala 460
1237 CAGATGAGACTCATTGAGAAACGGCAGAT-----CAGCAGGAAGCT 1278
|||.....|
Db 461 GluGluAspLysLysThrGluIleLeuAspProAspThrLeuGluAsnGlnAlaThrVal 480
|||.....|
1279 AAATGAGACCTCGAGATGCTGATGGCAGCAGGATTTACAGTATAGTACAACTAGAGCAG 1338
|||.....|
Db 481 GluPheHisSerGlyAspLysThr---MetGluCysGluLysLeuGlyLeuSerLysHis 499
|||.....|
1339 AAGCTTAGCGAA-----GACCAAGGGATAAAGGTAGAAATGAGAAAGCT 1383
|||.....|
Db 500 ThrThrAsnAspArgThrLysTyrlleAspAspThrValLysHisLysValLysIleLeu 519
|||.....|
1384 GCAATCCAGTCCGCAAGAGAACTCAGATCTTCAGCCTGTGATAGGGCCCTGCT 1443
|||.....|
Db 520 LysArgGluSerGlyGluGlyArg-----527
1444 GCCTCAAAATGATTTGCGCCCGGCTGTCTCAGCTGGCGCAGCCGACTCGGTACTGC 1503
|||.....|
Db 528 -----AsnSerSerAspCys 532
1504 AGTAATGACTGTATCTCAACACGCGCGCAGCACTAGTGAAGTTTCTAGCTCAGTAA 1563
|||.....|
Db 533 ArgAspArgGluIleLysLysTrpGlnLeuAlaProLeuArgLysMetGlyGlnProVal 552
|||.....|
1564 GAACAGAGGCAAGCCTAAAGAAAGATGAAGATGAAGCAGCAGAGCCAGCTCTCCG 1623
|||.....|
Db 553 LeuProArgArgSerSerGluGluLysSerGluLysIleProLysGluSerThrThrVal 572
1624 AATGCGGTGCTCAGGAGGTATTAATAATCTTCTGTGCACAGAGACCGCTCCAGAA 1683
|||.....|
Db 573 ThrCys-----ThrGlyGluLysAlaSerLysProGlyThrHisGluLysGlnGlu 599
|||.....|
1684 AAAAAAGAGACCAAGTGAAGAGGCGAGTGTGTGCTCCCTCGCGGAGTGAAGCACTCGGG 1743
|||.....|
Db 590 MetLysLysLysLysValGluLysGlyVal-----LeuAsn 601
1744 AAGGAAGCAGCTTGTGAGAGCAGCAGCCCTCG 1776
|||.....|
Db 602 ValHisProAlaAlaSerAlaSerLysProSer 612
RESULT 14
AAW79892
ID AAW79892 standard; protein; 1766 AA.
XX
AC
XX
AC
XX
06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3538.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US004098.

```

```

XX 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI: 2001-476283/51.
DR N-PSDB; AAK53025.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
in diagnosis and gene therapy.
XX
PS Claim 20; Page 376-377; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
(AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the
sequence listing were missing at the time of publication
XX
SQ Sequence 1766 AA;
Alignment Scores:
Pred. No.: 5,628-09 Length: 1766
Score: 232.00 Matches: 94
Percent Similarity: 36.43% Conservative: 63
Best Local Similarity: 21.81% Mismatches: 154
Query Match: 4.95% Indels: 120
DB: Gaps: 17
US-09-787-016A-1 (1-2610) x AAW79892 (1-1766)
QY 538 TCCCTGGAGAT---TCTGTGAGCCCGCTCTCTCCCGCCACAGACCCGAGACAGCC 594
|||.....|
Db 322 SerLeuSerAspLysSerHisAlaHisProGlyCysLeuLysGluProHisProAla 341
595 TCGAGGGCAGCGTGAAGAGCGCTTCTGAGACAGA---AGCGGCCCCAGTCTGTTCC 651
|||.....|
Db 342 GlnThrGlyHisValSerHisSerGlnLysGlnCysHisLysProGlnGlnAla 361
652 ACAGCTGTGAAGAAACGACACAGCTCTTCTGAAAAGGTGAAA---GGAGGGATGACCCAC 708
|||.....|
Db 362 ProAlaMetLysThrAsnSerHisValLysGluGluLeuGluHisProGlyValGluHis 381
709 GATGACACCTCCGATGATGACGCGATGCGCTGACCTTGAAGAG---CTT 756
|||.....|
Db 382 -----PheLysGluGluAspLysLeuLysLysLysProGlyLysAsnLeu 397
757 CAGATCGCTTCGCGAGGAAGCGGGAACAGGAGCCCTCAGAGGCCCTGAAAGGGATC 816
|||.....|
Db 398 GlnProArgGlnArgArgSerSerLysSerPheSerLeuAspGluProLeuPheIle 417
817 CAGAGTCGCTCGGAGAGAGCGCCCGGAGAGGGTCCCGCGAGAGACTGTGGCTCCGAG 876
|||.....|
Db 418 ProAspAsnIleAlaThrIleArgGlu-----GlySer--- 429

```


QY 877 GCCAGTGCACACTGTGGAGGCGTCTCTGCCCAAGTAAAGCAGGAGCCGAGACGATCAGGGG 936
Db 429 -----
QY 937 GTTGTGTCCTCCAGCTGGGAAAGATGACAGAGAGTAAGTTGAGGAGGAAGGCGGCTCAG 996
Db 430 -----
QY 997 GACATCAAGATGAGGAGCCTGGAGACTGTGGCGCGACCGAAGCCTGAATGTGAGGGTTAC 1056
Db 440 -----
QY 1057 GACCCCAACCCCTGTATTCGATTTGCGCCAGCCTCAACAACAGAGTTTATGATTGC 1116
Db 450 -----
QY 1117 TGTGACCGCTGTGAAGATGTTTTCATGGCGATTGTGTGGCCATTCTGAGGCTCGAGGG 1176
Db 462 CysGlyArgCysAspAspThrPheHisGlyAspCysValGlyLeuSerLeuSerGlnAla 481
QY 1177 AGCTTTTGAAGGAAGTGGGAGACTATATCTGCCAAACTGCACCATTTCTGCAAGTG 1236
Db 482 GlnGlnMetGlyGluGluAspLysGluThrValCysValLysCysCys-----Ala 498
QY 1237 CAGATGAGACTCATTTCAGAACCGCAGAT-----CAGCAGGAGACT 1278
Db 499 GluGluAspLysLysThrGluLeuLeuAspProAspThrLeuGluAsnGlnAlaThrVal 518
QY 1279 AAATGGAGACTGGAGATGCTGTATGGCACCGATTGTACAAATGATAGAACAAATAGACGAG 1338
Db 519 GluPheHisSerGlyAspLysThr--MetGluCysGluLysLeuGlyLeuSerLysHis 537
QY 1339 AAGTCTAGCGAA-----GACCAAGGATTAAGGTAGAAATGAGAAAGCT 1383
Db 538 ThrThrAsnAspArgThrLysThrLysLeuAspThrValLysHisValLysIleLeu 557
QY 1384 GCAAAATCCAAAGTGCAGAGAAACTCAAGATCTTCCAGCCTGTGTAGAGGCGCCTGGT 1443
Db 558 LysArgGluSerGlyGluGlyArg----- 565
QY 1444 GCCTCAAAATGTATGGCCCGCGGTGTGTACGTGGCGGAGCCGCGACTCGGTGTACTGC 1503
Db 566 -----
QY 1504 AGTAATGACTGTATCTCAAAACCGCGCAGCAGCAATGAATGCTTCAAGCTCAGGTAAA 1563
Db 571 ArgAspAsnGluIleLysLysThrGlnLeuAlaProLeuArgLysMetGlyGlnProVal 590
QY 1564 GACAGAGAACCAAGCCTAAAGAAAGATGAAGATGAAGCAGAGAGCCAGCTTCCG 1623
Db 591 LeuProArgArgSerSerGluGlySerGluLysIleProLysGluSerThrThrVal 610
QY 1624 AATGCGGTCTCAGGCGAGTATTAATCTCTCTGTGCAAGAGACAGCTCCAGAA 1683
Db 611 ThrCys-----ThrGlyGluLysAlaSerLysProGlyThrHisGluLysGlnGlu 627
QY 1684 AAAAAGAGACACACAGAGAGAGGAGTGTGTCTCCCTCGCGGAGTGAAGACTCGGG 1743
Db 628 MetLysLysLysLysValGluLysGlyVal-----LeuAsn 639
QY 1744 AAGAGAGCAGCTGTGAGACAGCAGCCCTG 1776
Db 640 ValHisProAlaAlaSerAlaSerLysProSer 650
RESULT 15
AAM41020
ID AAM41020 standard; protein; 1766 AA.
XX
AC AAM41020;
XX
DT 22-OCT-2001 (first entry)
XX

DE Human polypeptide SEQ ID NO 5951.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukemia.
XX Homo sapiens.
OS
XX WO200153312-A1.
PN
XX 26-JUL-2001.
PD
XX 26-DEC-2000; 2000WO-US034263.
PF
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDE; AAI60176.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
XX Example 2; SEQ ID NO 5951; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification

XX SQ Sequence 1766 AA;

Alignment Scores:
Pred. No.: 5,62e-09 Length: 1766
Score: 232.00 Matches: 94
Percent Similarity: 36.43% Conservatives: 63
Best Local Similarity: 21.81% Mismatches: 154
Query Match: 4.95% Indels: 120
DB: Gaps: 17

US-09-787-016A-1 (1-2610) x AAM41020 (1-1766)

QY 538 TCCTGGAGGAT---TCTGTGAGCCCGCCAGCCGAGAGCCGAGAGCC 594
Db 322 SerLeuSerAspLysSerHisAlaHisProGlyCysLeuLysGluProHisProAla 341

